Beyond the Label: Regulatory Slack and Forum Shopping in the Pharmaceutical Industry*

Charu N. Gupta
UCLA Anderson School of Management

Jennifer Kao
UCLA Anderson School of Management

May 15, 2024

Abstract

We analyze the relationship between firms' ability to leverage regulatory slack and their market entry strategies. Using newly constructed genomic measures of disease market similarity, we systematically document evidence of forum shopping, whereby pharmaceutical firms seek the most lenient regulatory environment for approval when drugs have multiple potential therapeutic uses. Firms seek regulatory approval in smaller disease markets to lower the costs of regulation and rely on complementary, non-regulatory pathways—in the form of unapproved, "off-label" drug use—to expand demand. Our data allow us to characterize the degree to which new technologies can exploit such opportunities, shedding light on how firms navigate regulatory environments to speed the entry of new products to market.

Keywords: market entry; regulation; forum shopping; pharmaceuticals; off-label use; genetics

*charu.gupta@anderson.ucla.edu, jennifer.kao@anderson.ucla.edu. We thank Victor Bennett, Amitabh Chandra, Matt Grennan, Matt Higgins, Katya Khmelnitskaya, Josh Krieger, Daniel Levinthal, Alex Olssen, Olav Sorenson, Noam Yuchtman, and seminar and conference participants at the Academy of Management (AOM) Annual Meeting, American Society of Health Economists (ASHEcon) Annual Conference, Food & Drug Administration, SoCal Strategy and OT Workshop, UCLA Strategy Seminar, Utah Winter Strategy Conference, and Wharton Health Care Management Workshop for helpful comments. We acknowledge the Sorenson Center for Discovery and Innovation Studies at the David Eccles School of Business for access to data. Financial support for this research came from the Price Center for Entrepreneurship & Innovation at UCLA Anderson and the National Institute for Health Care Management Foundation. Sam Lin and Jordan Youhanaie provided outstanding research assistance. Any errors are our own.

1 Introduction

Important strategic decisions for firms involve which product markets to enter and when. In industries where entry is regulated—such as transportation, banking, and health care—these decisions become more complex as firms must also consider the costs and benefits of seeking regulatory approval (Joskow, 2005; Vives, 2019; Maini and Pammolli, 2023). For example, the pharmaceutical firm Janssen sought initial U.S. approval for its drug Remicade as a treatment for severe Crohn's disease, despite earlier success in clinical trials in a larger disease market, rheumatoid arthritis (Melsheimer et al., 2019).[1] Following its initial approval, Remicade's use broadened to include not only rheumatoid arthritis but also several other autoimmune diseases. While this pattern of first seeking approval in smaller disease markets before expanding to wider applications could reflect differences in firm resources or scientific opportunities, we investigate an alternative hypothesis: Firms engage in "forum shopping," the act of choosing the most favorable regulator, certification process, or jurisdiction through which to conduct firm activities. In the context of the pharmaceutical industry, firms seek the most lenient regulatory environment for initial approval when drugs have multiple potential therapeutic uses.

In this paper, we document evidence consistent with the idea that forum shopping shapes firms' market entry decisions: Despite focusing their clinical trial investments in large disease populations, firms prioritize regulatory approval in small disease populations. Pharmaceutical firms make these strategic decisions because they can leverage a type of regulatory slack permitting the use of approved drugs in diseases for which they have not yet been approved ("off-label" use). Relying on data from large-scale genomic sequencing efforts, we construct a novel index of disease similarity that allows us to measure a drug's potential off-label disease markets. In doing so, we extend the literature on forum shopping by empirically demonstrating the role and relevance of a technology's "shoppability"—i.e., the degree to which firms can choose more favorable regulatory environments for a given technology.[2] Using this index and exploiting variation across disease areas in their

[1] Remicade received initial approval as an "orphan" drug, meaning it is used to treat a rare disease affecting fewer than 200,000 Americans. Specifically, its initial label was for the treatment of moderate to severe Crohn's disease in patients who had not responded to conventional therapies and for the treatment of patients with fistulizing Crohn's disease (FDA, 2024b). Rheumatoid arthritis affects roughly 1.5 million people in the U.S. (NIH, 2024).

[2] Forum shopping has been studied empirically in companies' choices of bankruptcy filing jurisdictions (Ellias, 2018), states of incorporation (Bebchuk and Cohen, 2003; Cain and Davidoff, 2012), litigation courts (Cohen et al.,

regulatory costs, we provide a series of empirical tests supporting the view that such forum shopping is more likely for drugs with high off-label potential, or shoppability.

Firms face an inherent trade-off when forum shopping (Lerner and Tirole, 2006): A more stringent certification process reduces the probability of certification but makes buyers more likely to adopt the product. In pharmaceuticals, these certification processes vary across disease markets ("indications"), making the setting ideal for the study of forum shopping. As a precondition of initial regulatory approval, pharmaceutical firms must undertake a series of expensive, risky, and time-intensive clinical trials to prove their drugs' safety and efficacy to the U.S. Food and Drug Administration (FDA) before initial market entry (Adams and Brantner, 2006; Mullard, 2016; Wouters et al., 2020). A growing body of evidence suggests that obtaining approval for drugs targeting smaller disease indications may be less costly and risky than targeting larger indications because of the possibility of conducting smaller clinical trials with patients who are more likely to respond positively to treatment (Chandra et al., 2019; Michaeli et al., 2023). Since regulatory policy permits drugs approved for one condition to be used for other, unapproved off-label uses, firms can still access larger disease indications after any initial approval (Larkin et al., 2014; Bradford et al., 2018; Shapiro, 2018; Dubois et al., 2023; McKibbin, 2023; Tunçel, forthcoming). As an estimated 90% of approved drugs have multiple therapeutic uses, off-label drug use is a common practice across many diseases (Gelijns et al., 1998).

Notably, the potential for off-label use breaks the link between tougher regulatory processes and access to larger disease markets by offering an alternative means of expanding a drug's usage to those patient populations without obtaining specific approval for them. In the absence of off-label use, firms' ability to forum shop is limited. As a result, firms would simply weigh the costs of approval for a market against its benefits. Prior studies suggest that a pharmaceutical firm has incentives to seek initial approval in a drug's largest possible disease indication (Acemoglu and Linn, 2004; Blume-Kohout and Sood, 2013). Yet when firms have the option to choose from multiple certification environments, they will opt for the one with the least stringent (and therefore least costly) standards conditional on being able to obtain certification (Lerner and Tirole, 2006). In the presence of off-label use, the benefits of taking on the tougher regulatory processes associated with

2019; Sytch and Kim, 2021), and store locations (Holmes, 1998). A separate literature in banking has explored the related concept of "regulatory arbitrage" (e.g., Houston et al., 2012; Karolyi and Taboada, 2015; Buchak et al., 2018).

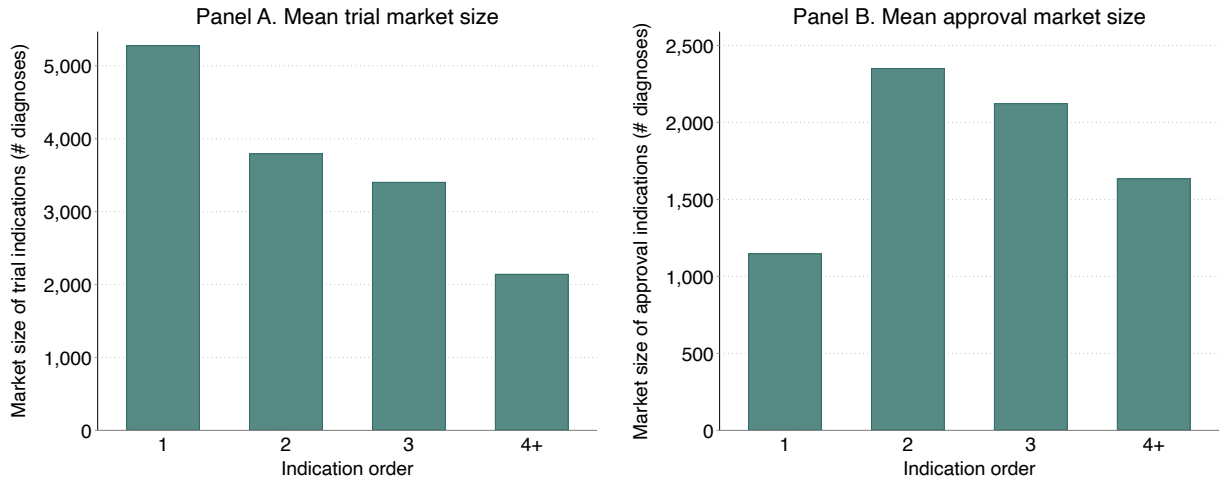
larger markets are reduced, and forum shopping becomes a useful strategy. As such, the current regulatory environment may inadvertently increase incentives for firms with products that have a large potential off-label market to prioritize regulatory approval of their drugs in small disease indications where they can more easily satisfy regulatory requirements. Such firms may then rely on complementary, non-regulatory investments (e.g., publication of trial results for unapproved disease indications in scientific publications) to expand demand to larger markets via off-label drug use.

The idea that forum shopping may shift firms' market entry strategies, while conceptually clear, is difficult to examine empirically. There are two main issues. First, variation in regulation requirements is rare, with regulatory changes typically anticipated and similar across product markets. As such, traditional empirical methods—which may rely on exogenous changes over time or across product markets—are ill-suited to assess how regulation may shift firms' approval decisions. Second, to understand how regulation shapes these decisions, it is necessary to identify the benchmark set of indications in which a firm has conducted clinical trials of its products and might plausibly seek regulatory approval. Researchers could then analyze the choice and timing of markets a firm pursues for formal approval against those in which it conducts trials but does not pursue formally. However, such potential markets are typically unobserved by researchers, making it difficult to fully characterize how firms alter their entry decisions in response to regulation.

Our empirical focus on drug development allows us to make progress on both of these issues.[3] To address the first issue, the challenge of identifying variation in regulation, we leverage differences in regulatory costs and off-label potential across disease indications. Relative to a "control" group of drugs with lower shoppability, those drugs with higher off-label potential are associated with more forum shopping: Their manufacturers can prioritize approval in disease indications with lower regulatory costs while relying on off-label drug use for entry to additional, larger markets without the significant investment needed for formal approval. To address the second issue, identifying a benchmark set of indications, we construct a comprehensive dataset matching drugs' FDA approvals with detailed clinical trial information. These data enable us to distinguish between two types of market entry investments: clinical trial investments and regulatory approval investments. With

[3]The pharmaceutical industry is projected to exceed $1.1 trillion by 2024, making it a noteworthy sector for analysis (IQVIA Institute, 2018).

FIGURE 1: MARKET ENTRY INVESTMENTS



NOTES: This figure shows the mean market size by indication order for both clinical trial investments (Panel A) and regulatory approval investments (Panel B) for cancer drugs approved from 1978 to 2016. Market size is measured by new diagnoses for an indication in the Surveillance, Epidemiology, and End Results (SEER) data. Trial indications are measured at the cancer site level and approval indications are at the cancer site-stage level.

these data, we compare the set and order of disease indications for which a firm tests a drug in clinical trials (the benchmark) against those in which it actually seeks approval.

To illustrate, Figure 1 explores entry investments across a drug's lifecycle, plotting mean market size by indication order—defined as the sequence of disease indications in which a given drug is clinically tested or approved—for a sample of cancer drugs approved from 1978 to 2016. Panel A indicates that firms prioritize large indications for their clinical trial testing, consistent with incentives to gather information about larger market sizes. However, Panel B indicates that firms prioritize smaller indications for regulatory approval. While such patterns could be due to differences in scientific opportunities (Krieger, 2021), market conditions (Acemoglu and Linn, 2004), or intellectual property (IP) protection (Budish et al., 2015), the indication order-market size relationships persist even after controlling for these factors. Rather, the trends are consistent with the view that firms forum shop, seeking initial regulatory approval in smaller disease indications where it is less costly and risky.

Motivated by this finding, we present evidence from a series of empirical tests. First, we reexamine the indication order-market size relationship for the drugs' entry investments, accounting for each

drug's shoppable potential as measured by the market size associated with its off-label disease indications. One major challenge of this test is that we cannot directly measure the ex-ante potential for a drug's off-label use. Two features of the market for cancer medicines—the largest pharmaceutical market in terms of spending (IQVIA Institute, 2018)—allow us to address this challenge. First, among oncology drugs, multiple uses are common, and estimates of off-label use range from 50 to 75% (Pfister, 2012).[4] Second, the rise of cancer genome sequencing enables us to use genetic data to identify likely off-label disease indications. Cancer sequencing is an advance in medical technology which systematically catalogs the genetic aberrations underlying different types of cancer. By comparing the DNA sequences of cancer cells to those of normal tissue, researchers can characterize the genetic mutations likely driving the progression and growth of specific cancers and determine similarities across different cancer types (Weinstein et al., 2013). Based on sequencing data, we construct a novel index of disease similarity between cancer sites. For example, cancer mapping efforts have revealed the occurrence of the same genetic mutations underlying both ovarian and breast cancer (TCGA Research Network, 2011). This suggests that ovarian cancer may be an off-label site for a drug approved for breast cancer (Pleasance et al., 2022). Using this index, we can characterize the shoppability of each drug and approximate the total market size of its different indications, including any off-label use.

Accounting for the size of expected off-label markets, the indication-market size relationship reverses: We identify a strictly negative and significant relationship between the order of regulatory approval for an indication and its *total* market size. That is, while firms seek initial approval for smaller disease indications, these indications are associated with the largest potential off-label markets. Though we caution that our estimates are not causal, this evidence is consistent with forum shopping: Firms prioritize smaller approval markets with more lenient regulatory processes and rely on off-label use to enter new markets without formal approval.

To investigate settings in which the effects of forum shopping are most salient, we also examine heterogeneity across drug and firm types. We confirm that the strategy of initial approval in smaller disease indications before expanding into larger ones is most prevalent among drugs designated as

[4]This is due to favorable off-label reimbursement policies in cancer treatment, as well as other factors (e.g., high disease severity and fewer treatments for rare cancers) encouraging physicians to experiment beyond formally approved uses.

"orphan" drugs by the FDA. Drugs with an orphan designation benefit from regulatory incentives for treating diseases that affect relatively few people. Further, we find that larger firms, likely those with substantial resources to encourage off-label drug use, are more likely to prioritize regulatory approval of their drugs in smaller disease indications compared with smaller firms that may be limited by financial constraints.

Finally, we consider the managerial and policy implications of our findings. We show that firms' reliance on forum shopping has important implications for the quality and speed of their investments. Clinical trials associated with off-label indications are less likely to be costly, high-quality (i.e., randomized and controlled) trials. Additionally, in a back-of-the-envelope calculation, we find that pharmaceutical firms can enter the market 7.9 months more quickly by seeking a drug's initial regulatory approval in a small market relative to a large one, translating into $99.6 million in value from clinical trial savings and revenues over this period. The size of the cost savings and our findings on the prevalence of this strategy should prompt regulators to consider whether firms are actively avoiding engaging in regulatory processes and the implications for consumers. Currently, numerous drugs are recommended and used off-label for important health conditions—for example, the use of aspirin prophylaxis for coronary disease in certain high-risk patient populations—yet off-label drug use without sufficient evidence is also associated with higher rates of adverse events (Wittich et al., 2012; Eguale et al., 2016; Richardson, 2016).

Across a variety of industries, the belief that entry regulation may shape firms' market entry decisions has fueled considerable policy attention. For example, in the transportation industry, ride-sharing apps such as Uber and Lyft have been able to bypass city taxicab regulations by arguing they are technology platforms and not transportation providers (Posen, 2015). Similarly, in the financial industry, financial technology ("fintech") companies like PayPal refrain from certain activities, such as holding customer funds or making loans, to avoid being regulated as banks (Douglas, 2016; Vives, 2019). Regulators must strike a balance between expediting consumer access to new products or services and ensuring their quality via rigorous standards and examination.

The paper proceeds as follows. Section 2 describes the institutional background behind pharmaceutical entry regulation and off-label drug use in the United States. Section 3 outlines the data, including the construction of our novel disease similarity index, and provides summary statistics.

Section 4 gives the empirical results, and Section 5 discusses managerial and policy implications of forum shopping. Section 6 concludes.

2 Institutional Background

2.1 Pharmaceutical Entry Regulation

The drug development process for a novel drug typically begins with extensive preclinical laboratory research that involves testing the candidate on animals and human cells. Once complete, the manufacturer submits an Investigational New Drug (IND) application to the FDA and begins the most expensive aspect of drug development: gathering safety and efficacy evidence from a series of clinical trials in which costs increase with each subsequent phase. Phase I trials must demonstrate a drug's safety and optimal dosage in healthy volunteers before the testing moves to phase II, in which safety and efficacy are tested in 100 to 300 patients. If successful, the drugs move to phase III trials, in which safety and efficacy are tested in 300 to 3000 patients (FDA, 2024a). Upon successfully completing phase III trials, the sponsor will submit a New Drug Application (NDA) to the FDA for final approval. The entire process is long (often taking between 8 and 12 years), costly (typically costing a manufacturer between $300 million and $2.6 billion), and risky (only 9% of drugs that initiate clinical testing receive regulatory approval) (DiMasi, 2001; DiMasi et al., 2003; Adams and Brantner, 2006; CSDD, 2014; Danzon and Keuffel, 2014; Wouters et al., 2020).

The development and review process is indication-specific (i.e., a drug receives regulatory approval to treat a specific disease in a specific population). Consequently, to expand a drug's label to acquire approval for a new use, the manufacturer must submit a new IND, undertake safety and efficacy trials, and submit a supplemental New Drug Application (sNDA). The amount of resources involved depends on the similarity between the original and new indications (FDA, 1998a). For example, if the original and new uses are closely related, manufacturers seeking approval for the new use may skip phase I trials and conduct fewer phase II trials (FDA, 1998a).[5] While some drug uses are discovered after the drug is first approved, others are known prior to approval. Indeed, industry reports highlight that "indication sequencing"—which includes identifying the choice of a drug's "lead indication"—is an important decision for firms (IQVIA, 2023).

[5] Examples include a new stage of the same disease or the same disease in a new population.

For manufacturers, the cost of satisfying regulatory requirements varies across indications, leading to the potential for forum shopping when making market entry decisions. To the best of our knowledge, there are no publicly available data estimates of average drug development costs by indication. However, some quantitative evidence suggests that drug development may be less costly for smaller indications relative to larger indications: Participant recruitment and enrollment costs constitute a major portion of trial costs (Sertkaya et al., 2016), and trials for smaller indications often require fewer participants than those for larger indications (Hee et al., 2017). While identifying suitable trial participants for trials in smaller indications may be challenging, firms have increasingly utilized genetic sequencing and biomarkers to target specific patient subgroups that are most responsive to treatment, which may reduce costs and risks (Chandra et al., 2019; Michaeli et al., 2023).

Consistent with this trend, since the 1983 passage of the Orphan Drug Act of 1983 (ODA), which offers incentives such as tax benefits and regulatory exclusivities to firms developing drugs for orphan diseases (i.e., rare diseases affecting less than 200,000 individuals), there has been a significant increase in the number of approved orphan drugs (from 10 in 1990 to 77 in 2017) (Miller and Lanthier, 2018).[6] On average, drugs treating orphan diseases experience a shorter regulatory review period than those for non-orphan diseases (Seoane-Vazquez et al., 2008). Many of these orphan drugs are subsequently used to treat common conditions (Miller and Lanthier, 2018). Additionally, Online Appendix Figure A1 shows that drugs tested in smaller indications are associated with higher approval rates. Taken together, the possibility of lower regulatory costs and a quicker approval process for smaller disease indications, combined with the potential for off-label drug use, leads to forum shopping. By prioritizing smaller disease indications when seeking initial regulatory approval, manufacturers may be able to launch their drugs earlier.

2.2 Off-Label Drug Use

Rather than expanding demand for its drug by seeking supplemental approvals, a manufacturer may choose to exploit regulatory slack in the form of off-label drug use. The practice of using approved health care technologies for unapproved uses is legal and common, particularly among treatments for cancer, cardiovascular diseases, and psychiatric diseases (Stafford, 2008). Off-label use occurs

[6] For more details, see `https://www.fda.gov/industry/medical-products-rare-diseases-and-conditions/designating-orphan-product-drugs-and-biological-products`.

for various reasons, including the lack of FDA-approved therapies for certain populations and the practice of using the same treatment for different but similar conditions (Wittich et al., 2012). For example, off-label use of psychiatric medicines is common in children because children are rarely included in clinical trials for drug approval (Lee et al., 2012). Many mental illnesses share the same or similar symptoms, motivating physicians to use one drug approved for a particular condition to treat another.[7]

The FDA recognizes that off-label use can be clinically appropriate under some circumstances, but is concerned that widespread off-label drug use may lead to public health risks due to the lack of rigorous research supporting such use (FDA, 2014). As a result, the FDA prohibits the direct promotion of off-label uses to physicians and patients. Despite this, evidence suggests that manufacturers still generate clinical evidence as a means to encourage off-label use.[8] (Note that off-label drug use would still occur without supporting clinical evidence, driven instead, by physician experience and experimentation.) Additionally, the agency's policy forbidding off-label advertising has gradually loosened and been challenged over time (FDA, 2014). Manufacturers are currently permitted to respond to unsolicited questions about off-label uses from health care professionals and to disseminate information describing off-label uses from peer-reviewed journal articles, textbook chapters, and clinical practice guidelines (Avorn et al., 2015). As a result, rather than spend the significant resources necessary to obtain regulatory approval for supplemental indications, manufacturers may choose to rely on these complementary, non-regulatory investments to expand demand via off-label drug use.

3 Data

To empirically investigate whether pharmaceutical firms forum shop when making market entry decisions, we evaluate investments associated with drugs approved by the FDA. Our main analytic dataset consists of 129 cancer drugs first approved by the FDA between 1978 and 2016. For each

[7]The prevalence of off-label use depends on a physician's propensity to prescribe a drug with limited evidence of safety and efficacy. In practice, physicians who engage in off-label use are rarely accused of medical malpractice. The process of informed consent does not require physicians to disclose that a drug is being used off-label. Further, off-label use is not necessarily negligent if the off-label use is included in the current standard of practice.

[8]The U.S. Department of Justice has charged and fined several major companies with illegal off-label promotion, including Eli Lilly ($1.4 billion in 2009), Pfizer ($2.3 billion in 2009), GlaxoSmithKline ($3 billion in 2012), and Abbott ($1.6 billion in 2012).

cancer drug, we obtain its NDAs and sNDAs from the Clarivate Cortellis Competitive Intelligence Global database and the FDA's Drugs@FDA database. We categorize approved indications at the cancer site-stage level (e.g., "ovarian-metastatic").[9] We then match these data on regulatory approval investments to data on clinical trial investments, which come from the Clarivate Cortellis Clinical Trials Intelligence database. Data limitations prevent us from categorizing clinical trial indications to the cancer stage level; as a result, our clinical trial analyses are conducted at the drug-cancer site level.

As a proxy for each indication's market size, we collect data on the number of new diagnoses associated with each cancer's site and stage (Budish et al., 2015). These data come from the Surveillance, Epidemiology, and End Results (SEER) database, available from the National Cancer Institute (NCI). We focus on five-year lagged averages of market size, where we calculate these lags relative to either the indication approval (for regulatory approval analyses) or trial start year (for clinical trial analyses).

3.1 Cancer Genome Mapping Data

We construct a novel proxy for each product's shoppable potential: the drug's expected off-label market size. To do so, we rely on cancer genome sequencing—an advance in medical technology which systematically catalogs the genetic mutations underlying different cancer types. Genetic sequencing compares the DNA sequences of cancer cells to those of normal tissue, so that researchers can determine the mutations behind specific cancers and understand similarities across different cancer types (Weinstein et al., 2013). Using genetic sequencing data, we characterize the similarity between different diseases to define a drug's expected off-label cancer sites. For example, cancer mapping efforts have shown that if a cancer site (e.g., breast cancer) shares genetic mutations (e.g., BRCA1) with another cancer site (e.g., ovarian cancer), drugs approved for one site may be used off-label to treat the other (TCGA Research Network, 2011; Pleasance et al., 2022). This approach is similar in spirit to research conducted by the bioinformatics community using genomic data to

[9] Note that FDA-approved indications are typically more granular than the cancer site-stage, and a drug can receive multiple approvals for the same cancer site-stage. For instance, letrozole was originally approved for "advanced breast cancer in postmenopausal women with disease progression following anti-estrogen therapy." It was later approved for "first-line treatment of postmenopausal women with hormone receptor-positive or hormone receptor-unknown locally advanced or metastatic breast cancer." In both cases, the approval was for the cancer site "breast" and the cancer stage "metastatic."

aid drug development efforts (Cheng et al., 2019; Tanoli et al., 2021). This includes leveraging data from disease-specific sequencing efforts (e.g., the Cancer Genome Atlas) and genome-wide association studies (GWAS) to uncover biological linkages across diseases that are then used to identify new uses for existing drugs (Uffelmann et al., 2021).

We obtain information on gene-cancer pairings uncovered in large-scale cancer mapping efforts from the publicly accessible Catalogue of Somatic Mutations in Cancer (COSMIC) Cancer Gene Census (CGC) database (Sondka et al., 2018; Tate et al., 2018).[10] The COSMIC team curates cancer genome data from hundreds of genetic sequencing studies and literature to catalog the set of genes containing mutations that are causally associated with cancer. In the CGC, each gene (e.g., BRCA2) is linked to the set of cancers (e.g., breast cancer, ovarian cancer) where mutations in that gene are likely contributors to the disease's development.[11]

3.1.1 Gene-Based Disease Similarity

Using the CGC data, we then estimate the similarity between cancer sites using the extent of overlap between the sets of genetic mutations associated with each cancer type. In the spirit of Krieger et al. (2022), we quantify the similarity between two different cancer sites by calculating the Tanimoto distance (Jaccard coefficient). This measure calculates the distance between each set of genetic mutations associated with each of the cancer sites. For example, the similarity index s between cancer sites A and B is the intersection of A and B's genetic mutations divided by the union of these mutations:

$$s_{A,B} \equiv \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|} \tag{1}$$

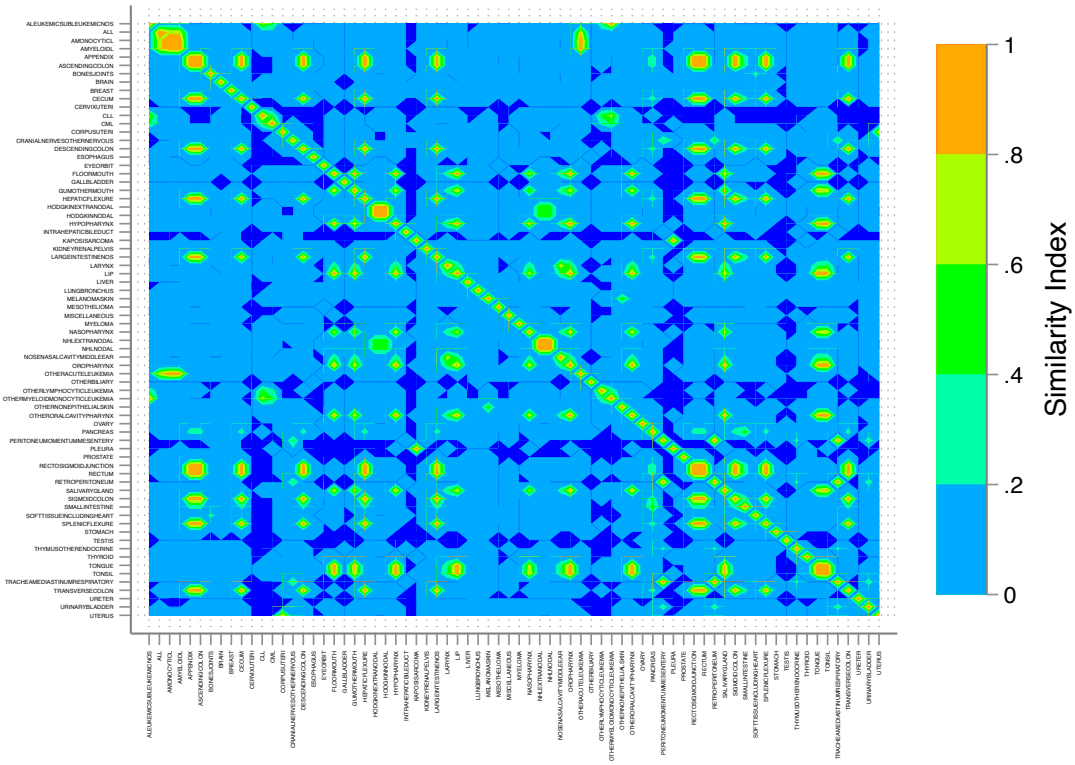
A similarity index of 0 implies that a pair of cancer sites are not closely related and have no common mutations, while a measure of 1 implies that they are closely related and have exactly the same set of mutations. Figure 2 provides a heat map of our similarity index across all 80 potential cancer sites. Among different cancer sites, the mean index is 0.074 (SD = 0.198).[12]

[10] For more details, see `https://cancer.sanger.ac.uk/census`.

[11] We consider both somatic and germline mutations. Somatic mutations occur in any cells of the body except the germ cells (sperm and egg) and are thus not passed on from parents to children. Germline mutations are changes in DNA inherited from parents.

[12] As an example, Online Appendix B provides the mutations for breast and ovarian cancer, including their overlapping mutations (Table B1), and shows the calculation of their similarity index. Breast and ovarian cancer have a similarity index of 0.118.

FIGURE 2: SIMILARITY INDEX HEAT MAP



NOTES: This figure shows a heat map of our similarity index across all cancer sites in the COSMIC Cancer Gene Census (CGC). Cancer site pairs associated with lighter colors are more similar.

3.1.2 Calculating a Product's Shoppability: Expected Off-Label Market Sizes

We use the similarity index to determine the expected off-label market size for each indication. This calculation is based on the market size for cancer sites related to indication i and the corresponding similarity indices:

$$\textit{Off-Label Market Size}_{i,t} = \sum_{j \neq i} s_{i,j} \textit{Market Size}_{j,t} \tag{2}$$

Here, indication i has an off-label market size at time t that is defined as the weighted sum of the market sizes of all other cancer sites j ($j \neq i$) at time t, where the weight is the similarity index between cancer sites i and j. To illustrate, consider a simple example between three cancer sites, A, B, and C. Assume that cancer site B has a similarity with cancer site A ($s_{A,B} = 0.09$) and a similarity with cancer site C ($s_{B,C} = 0.01$). Suppose that at time t, the market size for cancer site

A is 100 and the market size for cancer site C is 20. At time t, the off-label market size for a drug initially approved in cancer site B is $(0.09 \times 100) + (0.01 \times 20) = 9.2$.

Before continuing, we note the limitations to this measure. First, this is likely to be an underestimate as off-label drug use may not be supported by scientific evidence (Radley et al., 2006) and can be driven by other factors, such as costs and drug side effects (Stafford, 2008). In a following section, we address this by including a robustness check where we directly incorporate *actual* measures of off-label drug use. Second, the evidence from the CGC may not fully capture the true level of overlap across cancer sites: The creators of the CGC describe it as being a "conservative but high-confidence list" of genes associated with cancer, raising questions about the possibility of false negatives when deciding which gene-cancer associations to include in the database.[13] As a robustness check, we also construct a less restrictive disease similarity measure based on a broader range of gene-cancer associations identified in large-scale cancer sequencing studies. As we show in Section 4.2, the results using this less restrictive disease similarity measure are largely the same.

3.2 Additional Data

We construct several controls. As a proxy for the level of competition faced by each drug in a given indication, we count the cumulative number of drug approvals in the same indication in the year prior to the indication approval year (for regulatory approval analyses) or the trial start year (for clinical trial analyses). To capture regulatory incentives associated with subsequent market entry investments, we compile data on whether the drug ever received an orphan drug designation. Finally, we incorporate information on IP protection using data from the FDA's Orange Book and the United States Patent and Trademark Office (USPTO).[14] We create two controls for IP protection: *Primary IP protection* measures the months from each indication approval (or trial start) date to when the drug's primary IP expires and is generally considered the strongest form of IP protection, with almost certain enforcement.[15] We consider the primary IP expiration to be the latter of either the molecule patent or the new chemical entity exclusivity expiration. *Potential IP protection* gives

[13] The CGC's cancer experts apply strict criteria when determining the set of gene-cancer associations to include in the database. For more information, see `https://www.sanger.ac.uk/data/cancer-gene-census/`.

[14] Drugs are protected by two types of IP rights: patents granted by the USPTO and regulatory exclusivities granted by the FDA.

[15] For those trials taking place before a drug's initial launch, we assume the firms have an ex-ante expectation of what this primary IP expiration will be. For trials or approvals taking place after the primary IP has expired, we consider this measure to be zero.

TABLE 1: SUMMARY STATISTICS

	Mean	SD	Min	Max
Drug level				
Number of Unique Approval Indications	4	6	1	35
Number of Unique Trial Indications	46.7	23.6	1	78
Share with Orphan Disease Designation	0.64	0.48	0	1
Drug-indication level				
Market Size: Approval Indications (Diagnoses)	985	1,749	2.4	9,104
Market Size: Trial Indications (Diagnoses)	1,616	3,218	4.6	17,915
Market Size: Potential Off-Label (Diagnoses)	11,084	13,130	0	46,010
IP Protection: Primary (Months)	123.5	55.3	0	277
IP Protection: Potential (Months)	173.9	50.6	44	346
Competition: Approval Indications (Total Approvals)	3.1	3.6	0	20
Competition: Trial Indications (Total Approvals)	3.3	3.8	0	21

NOTES: This table shows summary statistics for our dataset of cancer drugs approved from 1978 to 2016. Statistics are at the cancer site level for trial indications and the cancer site-stage level for approval indications and potential off-label market size. Market size is measured by new diagnoses for an indication in the SEER data, while off-label market size is measured by the proportion of new diagnoses for any potential off-label indications, with the proportions given by our disease similarity index.

the maximum possible term of monopoly protection, if all IP on a drug were upheld. It measures the months from each indication approval (or trial start) to when the final IP on the drug expires.[16]

Table 1 presents summary statistics for our sample of 129 oncology drugs approved between 1978 and 2016. The average drug was tested in 47 different trial indications (cancer sites) and received FDA approval in 4 unique approval indications (cancer site-stage observations).[17] Across the sample, 64% of drugs have received an orphan drug designation. Across all drug-indications, mean primary IP protection remaining at the time of approval was 124 months (10.3 years) and mean potential IP protection was 174 months (14.5 years). The maximum primary IP protection remaining is 277 months (23.1 years) and the maximum potential IP protection is 346 months (28.8 years). Mean off-label market size (11,084 diagnoses) is more than 11 times the mean market size of approval indications (985 diagnoses).

[16] For trials occurring before a drug's initial launch, we consider the patents and exclusivities in effect at launch to calculate this measure. For all trials after launch, we consider the patents and exclusivities in effect at the trial start date.

[17] Online Appendix Figure C1 shows the complete distribution of trial and approval indications per drug.

4 Empirical Results

To examine how forum shopping shapes firms' decisions regarding which indications to enter and when, we document evidence from a series of empirical tests. Our goal is not to provide causal evidence of the effects of entry regulation, but rather to descriptively document the role and relevance of forum shopping in shaping firms' market entry decisions. First, we show that despite prioritizing clinical trials in larger indications, firms prioritize regulatory approval in smaller indications. This relationship persists even after controlling for detailed drug and market characteristics, suggesting the potential for forum shopping. In our second empirical test, we directly account for a drug's shoppable potential as proxied by its expected off-label market. Our findings are consistent with the view that firms prioritize smaller approval markets and rely on off-label use to enter new markets without formal approval. Finally, we exploit heterogeneity by drug and firm type to examine settings in which forum shopping is more prevalent.

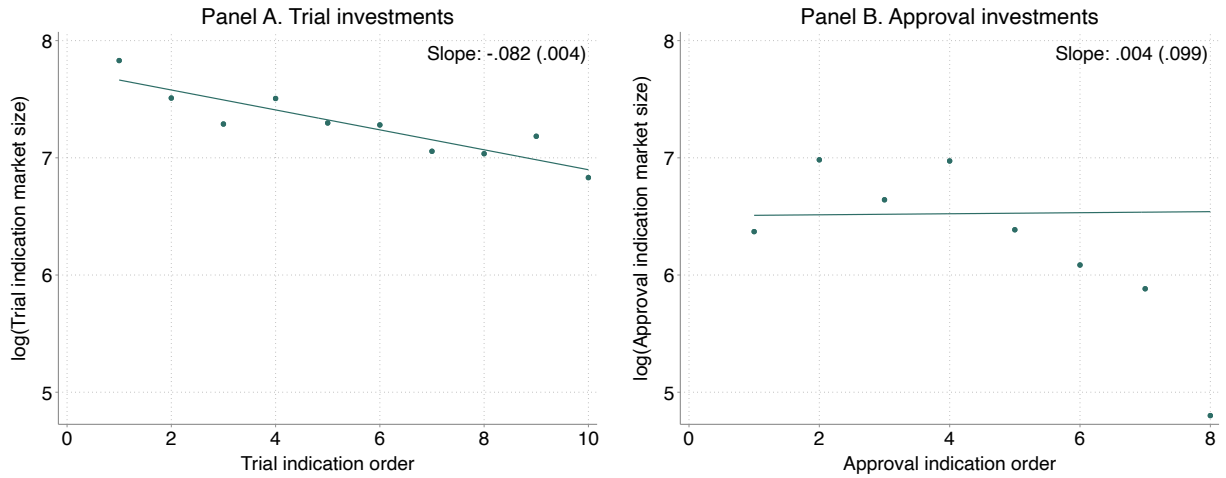
4.1 Analysis of Market Entry Investments

We aggregate our data on market entry investments, market size, and drug characteristics into drug-indication order (e.g., letrozole - first indication) observations, where indication order is defined as the sequence of indications in which a given drug is clinically tested (or approved).[18] Figure 3 explores trends in market entry investments by plotting the relationship between indication order and market size for our sample of drugs. Consistent with Figure 1, Panels A and B illustrate key differences between the timing of clinical trial and regulatory approval investments. Panel A documents that trial indication order is strongly negatively correlated with market size; when it comes to clinical testing, firms target the drugs' largest markets first. In contrast, Panel B documents that approval indication order has no discernable relationship with market size; firms do not prioritize approvals in larger markets.

To understand the dichotomy between firms' clinical trial and regulatory approval investments, Table 2 formalizes the relationship between indication order and market size. For drug d and

[18] When a single indication order corresponds to multiple indications (e.g., a drug is initially clinically tested in both breast and ovarian cancers in trials starting on the same date), we take the average of their market sizes.

FIGURE 3: TRENDS IN MARKET ENTRY INVESTMENTS



NOTES: This figure shows trends in market entry investments for cancer drugs approved from 1978 to 2016. The level of observation is the drug-indication order. Panel A shows the relationship between trial indication order and market size; the number of observations is 1,980. Panel B shows the relationship between approval indication order and market size; the number of observations is 187. Market size is measured by new diagnoses for an indication in the SEER data, and we use the log of the five-year average market size relative to either the trial start year (Panel A) or approval year (Panel B). Each marker represents binned averages for a given indication order. For ease of interpretation, we display up to the 10th trial indication.

indication order i, we estimate the following:

$$MarketSize_{d,i} = \alpha + \beta_1 IndicationOrder_{d,i} + \gamma X_{d,i} + \epsilon_{d,i} \tag{3}$$

Our outcome variable $MarketSize$ is the natural log of the lagged five-year average market size associated with indication order i for drug d. The coefficient on $IndicationOrder$ is our main estimate of interest. We investigate this relationship by conditioning on a series of controls X, including the initial approval year for the drug; its cancer group or type (e.g., stomach cancer and esophageal cancer would belong to the "digestive system" cancer group); competition (i.e., the cumulative number of drug approvals in the same indication as of the prior year); regulatory incentives (i.e., whether it has an orphan designation); and measures of intellectual property protection (i.e., its primary and potential IP). Each of these controls may influence the relationship between indication order and market size. For example, Budish et al. (2015) highlight that firms have reduced incentives to pursue research for early-stage cancers relative to late-stage ones because of the former's longer development times and shorter resulting patent terms post-launch. This has implications for our findings: If early-stage cancers have larger market sizes, the relationship

Table 2: Ordering of Market Entry Investments

	Clinical trial investments		Regulatory approval investments	
	(1)	(2)	(3)	(4)
Indication order	-0.0823***	-0.0471***	0.00448	-0.0265
	(0.00371)	(0.00576)	(0.0995)	(0.103)
Mean of Dep. Var.	6.829	7.040	6.513	6.522
Observations	1,980	1,523	187	182
Controls:				
Initial approval year	no	yes	no	yes
Indication group	no	yes	no	yes
Competition	no	yes	no	yes
Regulatory incentives	no	yes	no	yes
Intellectual property	no	yes	no	yes

Notes: This table shows the relationship between indication order and market size for cancer drugs approved from 1978 to 2016. The level of observation is the drug-indication order. Columns 1 and 2 examine clinical trial investments, and Columns 3 and 4 examine regulatory approval investments. The outcome variable is market size, measured by new diagnoses for an indication in the SEER data. We use the log of the five-year average market size associated with indication order. The number of observations differs between Columns 1 and 2 (and between Columns 3 and 4) due to missing data on control variables and dropped singletons. Robust standard errors are in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

between indication order and market size may be driven, in part, by patent-related factors. To account for this, we include detailed controls for each drug's patent and market exclusivity. All estimates are from ordinary least squares (OLS) models with robust standard errors.

Column 1 of Table 2 reports the raw correlation between trial indication order and market size. The estimated coefficient implies that a 1-unit increase in trial indication order is associated with an 8% ($\approx (exp(-0.0823) - 1) \times 100$) decrease in market size. Column 2 shows that this negative relationship between trial indication order and market size persists once all controls are included, and the relationship remains significant at the 1% level. Columns 3 and 4 repeat these same regressions for approvals instead of trials. Column 3 confirms that approval indication order has no significant relationship with market size. Once all controls are included in Column 4, the relationship remains quantitatively small and insignificant. We interpret this to mean that such factors as initial approval year, indication group, competition, regulatory incentives, and IP protection do not fully explain the gap between the timing of clinical trial versus regulatory approval investments and their relationship

to market size.[19] The results, however, are consistent with the view that given the relatively lower costs of obtaining regulatory approval for conditions with smaller market sizes and the potential for off-label markets, firms may forum shop by first seeking approval for smaller indications. We directly test this in the next set of analyses.

4.2 Incorporating Expected Off-Label Opportunities

Using the disease similarity index based on overlapping gene mutations described in Section 3, we consider in Table 3 *total* market size, including potential off-label markets, as our outcome of interest. For reference, Columns 1 and 3 repeat the specifications from Table 2, with focal indication market size as the outcome variable and including all controls, for trials and approvals, respectively. Column 2 then regresses total market size on indication order plus all controls for our trial sample. The relationship between indication order and market size remains negative, with an increase in indication order associated with a 6% ($\approx (exp(-0.0610) - 1) \times 100$) decrease in total market size. Column 4 repeats this exercise for our approval sample. In contrast to Column 3, where there is an insignificant relationship between indication order and focal market size, once we account for off-label markets the relationship becomes negative and statistically significant. An increase in indication order is associated with a 15% ($\approx (exp(-0.161) - 1) \times 100$) decline in total market size.[20] These results suggest that pharmaceutical firms still prioritize larger indications for clinical trial investments once we factor in potential off-label markets, but they forum shop by seeking initial approval for smaller (focal) indications, anticipating off-label expansion into others.

We probe the robustness of these results by using different measures of total market size that are generated with alternative similarity measures. These alternative measures are generated by directly using a broader range of gene-cancer associations identified in large-scale cancer sequencing studies. Following the bioinformatics literature, we focus on genetic mutations that occur at a high frequency within each mapping study, where we consider a genetic mutation as "high frequency" within a cancer site if it occurs in the top 10%, top 20%, or top 30% of the most frequently occurring

[19] Online Appendix Table D1 demonstrates the robustness of these results under an alternative specification where we use an indicator for the first indication as our main explanatory variable instead of indication order.

[20] Online Appendix Table D2 shows the robustness of these results using the alternative specification where an indicator for the first indication is the main explanatory variable.

Table 3: Ordering of Market Entry Investments, Incorporating Off-Label Opportunities

	Clinical trial investments		Regulatory approval investments	
	Focal market size (1)	Total market size (2)	Focal market size (3)	Total market size (4)
Indication order	-0.0471***	-0.0610***	-0.0265	-0.161*
	(0.00576)	(0.00608)	(0.103)	(0.0813)
Mean of dep. var	7.040	8.279	6.522	8.981
Observations	1,523	1,523	182	182
Controls:				
Initial approval year	yes	yes	yes	yes
Indication group	yes	yes	yes	yes
Competition	yes	yes	yes	yes
Regulatory incentives	yes	yes	yes	yes
Intellectual property	yes	yes	yes	yes

Notes: This table shows the relationship between indication order and market size for cancer drugs approved from 1978 to 2016. The level of observation is the drug-indication order. Columns 1 and 2 examine clinical trial investments, and Columns 3 and 4 examine regulatory approval investments. The outcome variable in Columns 1 and 3 is focal market size, while that in Columns 2 and 4 is total market size, including potential off-label markets. For both variables, we use the log of the five-year average market size associated with indication order. Focal market size is measured by new diagnoses for an indication in the SEER data. Total market size is measured by new diagnoses for the focal indication plus a proportion of new diagnoses for any potential off-label indications, with the proportions given by our disease similarity index. Robust standard errors are in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

mutations. Table 4 confirms that our results are largely unchanged when using these alternative similarity measures.

Table 4: Regulatory Approval Investments and Total Market Size, Using Alternative Similarity Measures

	All genes			Cancer Gene Census		
	Top 10% (1)	Top 20% (2)	Top 30% (3)	Top 10% (4)	Top 20% (5)	Top 30% (6)
Indication order	-0.203**	-0.215**	-0.223**	-0.206**	-0.233**	-0.236**
	(0.0943)	(0.0970)	(0.0971)	(0.0974)	(0.0984)	(0.102)
Mean of dep. var	10.049	10.199	10.246	10.283	10.561	10.650
Observations	182	182	182	182	182	182
Controls:						
Initial approval year	yes	yes	yes	yes	yes	yes
Indication group	yes	yes	yes	yes	yes	yes
Competition	yes	yes	yes	yes	yes	yes
Regulatory incentives	yes	yes	yes	yes	yes	yes
Intellectual property	yes	yes	yes	yes	yes	yes

Notes: This table shows the relationship between indication order for regulatory approval investments and total market size for cancer drugs approved from 1978 to 2016, using alternative similarity measures. Alternative measures are generated by directly using cancer genome sequencing data from 168 large-scale mapping studies. Genetic mutations are restricted to those that occur at a high frequency within each mapping study, where a genetic mutation is "high frequency" within a cancer if it occurs in the top 10% (Columns 1 and 4), top 20% (Columns 2 and 5), or top 30% (Columns 3 and 6) of most frequently occurring mutations. Columns 1 to 3 focus on genetic mutations occurring among all genes. Columns 4 to 6 focus on the set of genetic mutations occurring among genes found in the COSMIC Cancer Gene Census (CGC). The level of observation is the drug-indication order. The outcome variable is total market size, including potential off-label markets. Total market size is measured by new diagnoses for the focal indication plus a proportion of new diagnoses for any potential off-label indications, with the proportions given by our disease similarity index. We use the log of the 5-year average market size associated with indication order. Robust standard errors are in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

4.3 Heterogeneous Effects

In our previous empirical analyses, we show that the documented relationship between indication order and market size is robust to detailed controls for drug and market characteristics. The small sample size of our main cancer dataset limits our ability to meaningfully examine variation across different settings (e.g., various drug and firm types). As a result, we construct a dataset of approvals associated with a broad range of diseases. This "cross-disease dataset" consists of cancer and non-cancer drugs first approved between 1998 and 2021. We categorize approved indications using International Classification of Disease (ICD-9) codes. As a proxy for the market size of drugs in

this dataset, we use number of diagnoses from the Medical Expenditure Panel Survey (MEPS). We calculate yearly average market size at the ICD-9 level, using data from 1996 to 1997.

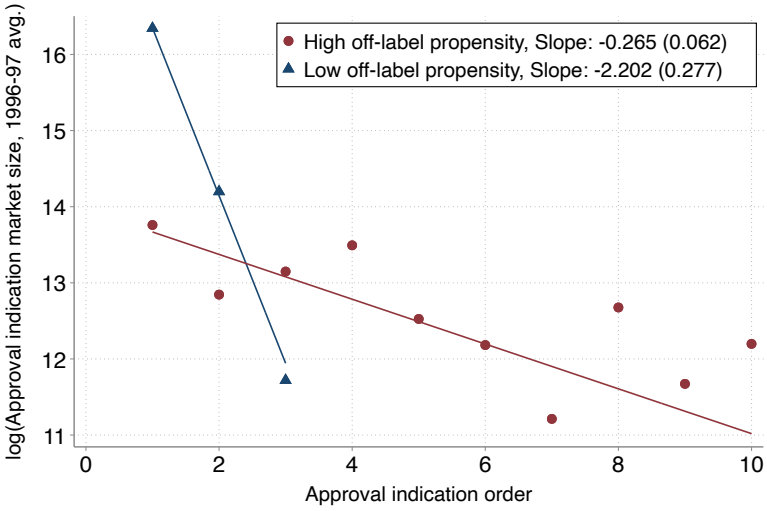
Before examining heterogeneity across drug and firm types, we probe the presence of forum shopping among drugs in this cross-disease dataset. In the previous analyses, we were able to approximate a drug's *expected* off-label market that is supported by scientific evidence by linking our cancer dataset to cancer genome mapping data. However, off-label drug use may occur in the absence of supporting scientific evidence (e.g., due to physician experience and experimentation). An alternative approach would be to rely on measures of *actual* off-label drug use. One challenge is that available measures of actual off-label drug use are generally restricted to survey and health insurance claims data that cover specific patient populations, limiting our ability to capture total off-label drug use. These limitations notwithstanding, we approximate a drug's actual off-label market size by using findings from Radley et al. (2006), which estimates off-label prescribing rates from a survey of physician prescribing practices. These estimates enable us to distinguish disease categories with high and low levels of shoppability, as measured by their off-label propensity.[21] Our resulting cross-disease sample covers 623 drugs, of which 490 drugs are initially approved in diseases with high off-label propensity and 133 drugs in diseases with low propensity.

Figure 4 shows that high off-label drugs have more approvals, on average, than low off-label drugs. Notably, the indication order-market size correlation is less negative among drugs first approved for diseases with a high off-label propensity relative to those first approved for diseases with low off-label propensity. On average, among the former, an increase in indication order is associated with a 23% ($\approx (\exp(-0.265) - 1) \times 100$) decrease in market size. Among the latter, an increase in indication order is associated with an 89% ($\approx (\exp(-2.202) - 1) \times 100$) decrease in market size. These findings are consistent with the view that unlike manufacturers of drugs in therapeutic areas with a high off-label propensity, those in therapeutic areas with a low off-label propensity do not have the option of expanding into other markets via off-label use, leading them to prioritize larger markets in their initial approvals.[22] Having established similar patterns across diseases as observed

[21] Diseases with low off-label propensity (low shoppability) include antidiabetics, antihypertensives, and antihyperlipidemics; those with high off-label propensity (high shoppability) include oncology, anticonvulsants, psychiatry, and antiasthmatics.

[22] We supplement these findings with several robustness checks. We find similar results after controlling for drug and market characteristics (Online Appendix Table D3); controlling for each drug's associated Anatomical Therapeutic Chemical (ATC) classification, a classification system by the World Health Organization (Online Appendix Table D4);

FIGURE 4: TRENDS IN REGULATORY APPROVAL INVESTMENTS ACROSS DISEASES, BY PROPENSITY FOR OFF-LABEL DRUG USE



NOTES: This figure shows the relationship between regulatory investments (FDA approvals) and market size for approved drugs across several disease categories from 1998 to 2021, by propensity for off-label drug use. The level of observation is the drug-indication order, where an indication corresponds to an ICD-9 code. Market size is measured by new diagnoses for an indication (ICD-9) in the MEPS data. We use the log of the market size associated with indication order averaged over 1996 and 1997. Each marker represents binned averages for a given indication order. For ease of interpretation, we display up to the 10th indication.

in our main analyses of cancer drugs, we now proceed with using this cross-disease dataset to explore potential heterogeneous effects.

4.3.1 Differences in Regulatory Incentives

As described earlier, the Orphan Drug Act (ODA) was designed to stimulate the development of drugs for rare diseases by offering incentives (e.g., tax credits, market exclusivities) to offset the low returns often associated with such therapies. Unrelated to these goals, growing anecdotal evidence suggests that firms engage in forum shopping by first seeking orphan drug designations and approvals for small indications before encouraging the off-label use of their drugs in more common diseases (Bagley et al., 2018). In this analysis, we categorize the cross-disease sample of drugs into two mutually exclusive groups: those with at least one orphan drug designation and those without any. Column 1 of Table 5 confirms a negative correlation between indication order and market size

and incorporating an alternative measure of actual off-label drug use (Online Appendix Table D5). We are grateful for the ATC-drug crosswalk provided by the authors of Kakani et al. (2022).

TABLE 5: HETEROGENEITY BY DRUG AND FIRM TYPE

	Orphan drug status		Firm size	
	Non-orphan drug (1)	Orphan drug (2)	Small firm (3)	Large firm (4)
Indication order	-0.529***	0.106*	-0.665***	-0.0726
	(0.195)	(0.0631)	(0.245)	(0.0657)
Mean of Dep. Var.	14.910	12.150	14.827	13.662
Observations	538	264	236	547
Controls:				
Initial approval year	yes	yes	yes	yes
Indication group	yes	yes	yes	yes

NOTES: This table shows the relationship between indication order for regulatory approval investments and market size for approved drugs across several disease categories from 1998 to 2021. The level of observation is the drug-indication order. The outcome variable is market size, measured by new diagnoses for an indication (ICD-9) in the MEPS data. We use the log of the average market size associated with focal indication order averaged over 1996 and 1997. Robust standard errors are in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

for drugs without orphan designations. In contrast, Column 2 shows that drugs with orphan drug designations prioritize approvals for smaller, less common indications before expanding to more widespread conditions. These findings confirm the intuition that firms may leverage regulatory opportunities to target small disease indications first before expanding into larger ones.

4.3.2 Differences in Firm Size

Firms that are able to credibly communicate the value of their drugs to consumers are likely to be more successful in expanding their off-label use. As a result, we would expect that larger, more experienced firms would be more likely to engage in forum shopping and to expand off-label drug use. To investigate, we split our drug sample into those manufactured by large firms (those publicly listed at the time of the drug's initial approval) and small firms. Consistent with this view, Columns 3 and 4 in Table 5 show that the negative indication order-market size relationship is less pronounced among large firms. This suggests that larger firms, with greater resources and capabilities, are more likely to engage in forum shopping.

4.4 Ruling Out Additional Alternative Explanations

We consider other possible explanations for the patterns we document. One is that firms may use expected trial length as a factor in their ordering of clinical trial indications. That is, if firms prioritize indications with longer anticipated clinical trials, that could be driving the negative indication order-market size relationship we observe for trials and the lack of a clear relationship for approvals. As a robustness check, we return to our main cancer dataset and consider trial indication order with respect to trial end dates rather than start dates. In Online Appendix Figure D1, we see the strong negative relationship between trial indication order and market size persists, reducing concerns that the gap between clinical trial and regulatory approval investments is due to trial length.

A second explanation could be that the non-negative indication order-market size relationship for regulatory investments reflects differences in FDA review timings rather than forum shopping by firms. If firms submit applications for larger indications first but the FDA requires longer review times for such applications, this would weaken our hypothesis that firms engage in forum shopping by seeking initial approvals in small markets. To address this possibility, we manually collected application submission dates for each approval from FDA review letters. In Online Appendix Figure D2, we thus carry out another robustness check where approval indication order is determined by submission dates rather than approval dates. We continue to see an insignificant relationship between approval indication order and market size, indicating that FDA review processes are not driving our results.

Finally, the FDA releases information only on successful approvals, raising concerns that the observed non-negative relationship between regulatory investments and indication order may primarily reflect FDA approval decisions rather than firms' forum shopping. While we cannot entirely rule out the influence of FDA preferences, our findings remain even after taking into account regulatory incentives. Further, firms are unlikely to submit applications to the FDA if they anticipate them to be rejected. Combined, these suggest that firms' strategic decisions do play a role in shaping their regulatory approval decisions.

5 Managerial and Policy Implications of Forum Shopping

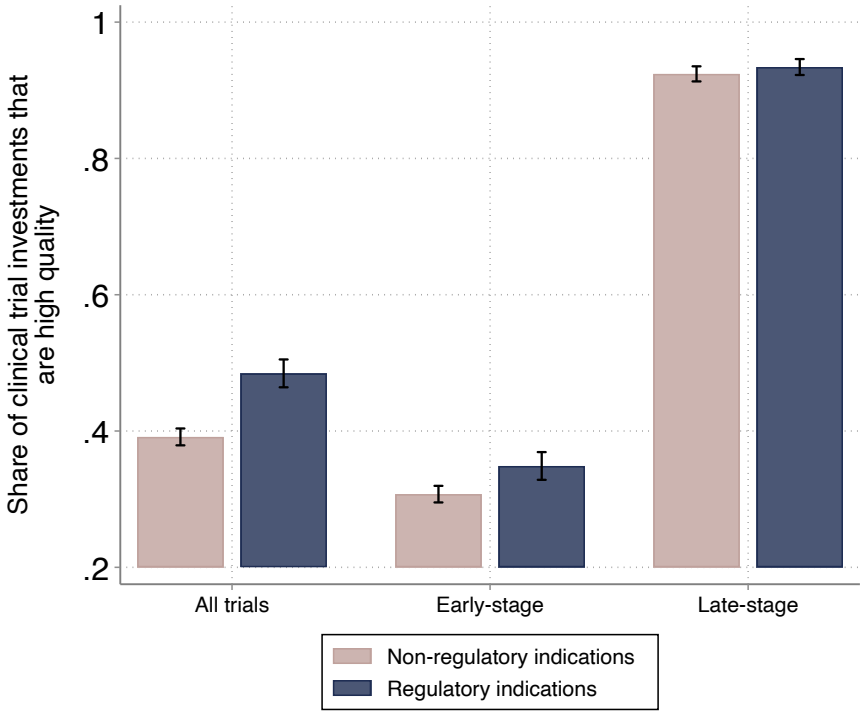
The paper's results indicate that firms forum shop for a more lenient regulatory environment by seeking a drug's initial approval in smaller markets and relying on off-label drug use to enter additional markets. In this section, we explore the implications of such forum shopping for firm managers on the design quality of trial investments, the speed of entry into new markets, and the number of markets they pursue. We also discuss the policy implications for regulators, who must balance the trade-off between expediting consumer access to new products and ensuring sufficient information about their quality.

5.1 Impact on Design Quality of Trial Investments

In this section, we examine the quality of non-regulatory trial investments, clinical trials with a primary purpose to expand off-label use (e.g., through dissemination of results in scientific publications) and not to support a future sNDA. A unique benefit of non-regulatory investments is that they, by definition, do not require the high costs associated with investments supporting regulatory approval. For example, trials aimed at expanding off-label use may not be randomized controlled trials, which are considered the gold standard (Bothwell et al., 2016). To investigate, we compare the quality of trials likely conducted for regulatory purposes with those likely conducted for non-regulatory purposes.

For each drug in our sample, we classify indications associated with clinical trial investments as either "non-regulatory" or "regulatory." A "non-regulatory indication" is one that is tested in clinical trials but for which the firm does not seek regulatory approval. A "regulatory indication" is one that receives FDA approval. Figure 5 compares the rate of trials for non-regulatory and regulatory indications that are high quality, with quality measured as the share of trials that are randomized and controlled. Looking at all trials, we find a significantly higher rate of high-quality trials for regulatory indications than for non-regulatory indications (48% vs 39%, respectively; $p < 0.01$). We then explore differences across early-stage (i.e., phases I and II) and late-stage (i.e., phases II/III and III) trials, which vary in length and cost, and find these differences persist, with a higher rate of high-quality early-stage trials among regulatory indications than among non-regulatory indications (35% vs 31%, respectively; $p < 0.01$). Consistent with the idea that firms are less likely to conduct

Figure 5: Quality of Clinical Trial Investments: Non-Regulatory vs. Regulatory Indications



Notes: This figure shows differences in the quality of clinical trial investments for non-regulatory versus regulatory indications for cancer drugs approved from 1978 to 2016. High-quality clinical trial investments refer to trials that are randomized and controlled. Bars give shares, while capped ranges provide 95% confidence intervals.

late-stage clinical trials that are high quality if they do not intend to pursue regulatory approval, we find little difference in trial quality among late-stage trials.

In addition to shedding light on how the ability to forum shop might shift the quality of firms' research investments, this exercise also mitigates concerns that the observed patterns are driven not by strategic rationales but by scientific ones (e.g., an sNDA is not sought since the drug is found to be unsafe or ineffective). If the difference in the pattern of clinical trial and regulatory approval investments were primarily due to scientific rationales, we would expect to see little difference in trial quality between indications with and without regulatory approval. Contrary to this expectation, we find statistically significant differences.

5.2 Impact on the Speed of Entry Into New Markets

Long development timelines are a major driver of high drug development costs (Wong et al., 2014). By forum shopping for initial regulatory approval in smaller indications, firms can potentially bring a drug to the market more quickly by conducting smaller clinical trials with a segment of participants more likely to respond to treatment (Chandra et al., 2019). To understand the benefit to firm managers of this strategy, we carry out a back-of-the-envelope calculation quantifying its dollar value.

We begin by comparing the speed with which cancer drugs enter the market when pursuing initial approval in small versus large indications. We restrict our sample to each drug's first approval and consider "small" indications as those within the lowest quartile of market size and "large" indications as the rest. Measuring from pivotal trial start date to approval date, drugs with small initial indications reach the market in 44.8 months on average, while those with large initial indications reach the market in 52.7 months, for a difference of 7.9 months. To translate this time savings into dollar savings, we make use of a recent study of clinical trial costs, which finds that each additional month in late-stage trials equals a median of $671,000 spent (Martin et al., 2017). Multiplying this figure by 7.9 months suggests that pharmaceutical firms can save more than $5.3 million in trial costs alone by forum shopping and seeking initial approval in a smaller market.

In addition to trial costs saved, firms also benefit from revenues obtained earlier. To determine per-drug revenues over this time, we turn to Schuhmacher et al. (2022), who examine new drugs launched and their total sales from 2011 to 2020. Given that we are considering drugs launched in small indications, we take a conservative approach by excluding blockbusters (with mean annual sales of >$1 billion) and high-selling ($0.5–0.999 billion) drugs from our revenue calculations. Looking only at low- (<$0.1 billion) and medium-selling ($0.1–0.499 billion) drugs, we determine these drugs have an average annual revenue of $143.2 million, equating to $94.3 million over 7.9 months.[23] We consider this to be a conservative estimate of total revenues because drugs initially launched in small indications may still generate large or even blockbuster-level sales, due to both the potential for off-label use and competitive factors allowing monopoly pricing to be high.

[23] We calculate annual revenues from Figure 1 of Schuhmacher et al. (2022) as follows: $82.1 billion total revenues for medium-selling drugs + $4.7 billion for low-selling drugs)/(51 medium-selling drugs + 50 low-selling drugs)/6 years average commercialization period over 2011–2020 = $143.2 million annual per-drug revenues.

Summing the costs saved from clinical trials and the revenues obtained via earlier market entry, we obtain a total value to firms of \$5.3 million + \$94.3 million = \$99.6 million per drug from forum shopping in a small market for initial approval.

5.3 Impact on the Number of Markets Entered

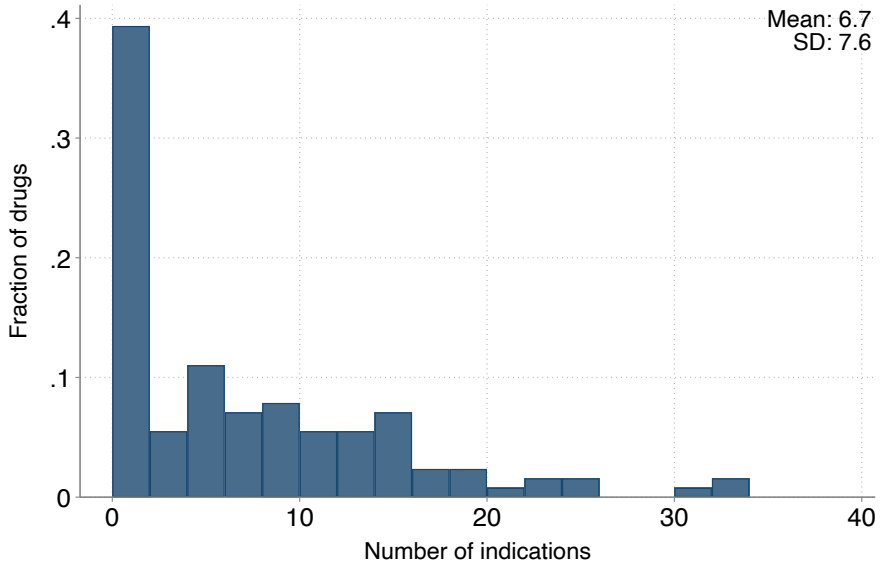
Next, we consider how the opportunity to lower regulated entry costs by forum shopping in smaller disease indications and leveraging off-label drug use impacts firm decisions on the number of markets to enter for a single drug. The high costs associated with regulatory approval for larger markets suggest that if off-label use were banned, we might expect firms to pursue regulatory approval for a portion of these indications, provided the benefits from approval in terms of market expansion exceed the costs of obtaining sufficient scientific evidence necessary for approval. For the remaining indications, for which the benefits of formal approval do not exceed the costs, these drug-indication pairs would become "missing" in the sense that firms would no longer pursue them. In light of this, a natural next question is, how many "additional" indications do firms explore because of regulatory slack in the industry that allows them to enter new markets through off-label drug use? Stated differently, how many off-label indications are associated with a drug?[24]

To investigate this, we examine the indications associated with clinical trial investments. We focus on (1) regulatory indications (which receive approval), and (2) the subset of non-regulatory indications likely pursued for off-label use. To construct the latter group, we note that the set of non-regulatory indications includes both those likely pursued for off-label drug use and those that were likely discontinued for scientific reasons. To identify the subset likely used for off-label drug use, we leverage their clinical trial design, as examined above in Section 5.2.

The process of identifying the set of off-label indications is as follows: We remove indications that were discontinued for scientific reasons using two methods. Our first is based on the idea that firms are likely to terminate trials early when they encounter safety or efficacy issues, making it unlikely that they will seek regulatory approval for the indication or disclose trial outcomes to facilitate off-label drug use (Cook et al., 2014). As a given drug will often have multiple trials (even within the same phase) studying the drug's use in a particular indication, we incorporate data on early

[24] For simplicity, this exercise does not consider the "size" of the associated indications.

FIGURE 6: DISTRIBUTION OF OFF-LABEL INDICATIONS



NOTES: This figure plots the distribution of off-label indications for cancer drugs approved from 1978 to 2016. Number of drugs is 127.

trial terminations and classify an indication as "discontinued" if it is ultimately unapproved and associated with a high (above-median) share of trial terminations. Our second method is motivated by the idea that firms conducting high-quality (i.e., randomized controlled) trials without seeking regulatory approval likely do so because of inadequate safety and efficacy findings. As a result, we assume that being unapproved and associated with a high (above-median) percentage of high-quality trials indicates the indication was likely also dropped for scientific reasons. "Off-label" indications encompass all non-regulatory indications that are not discontinued due to scientific reasons.

Using these data, we document the total number of regulatory and off-label indications associated with a drug. The average drug is associated with a total of 9.7 such indications, of which only 3 are approved and a mean of 6.7 are off-label; Figure 6 gives the full distribution of off-label indications across our sample.[25] Our findings support the idea that the opportunity for forum shopping likely increases the overall number of indications that firms focus on in their clinical trial efforts. A key caveat to interpreting these results is that we are examining equilibrium outcomes. If off-label drug use were banned, the total number of indications associated with trial investments might decline.

[25]Our main analyses consider trial indications at the site level and approval indications, for which we have more granular data, at the site-stage level. For this analysis of the number of markets, because we explicitly match trial to approval indications, all indications are at the site level. As such, we have a mean of 3 approval indications per drug at the site level versus, as indicated in Table 1, 4 approval indications per drug at the site-stage level.

However, it is not clear that the average number of indications for a given drug would fall to 3; firms might pursue regulatory approval for some indications initially intended for off-label use. Therefore, we consider the mean number of off-label indications (6.7) as an upper bound on the number of additional indications that firms might still pursue for a given drug if off-label use were prohibited.

5.4 Impact for Off-Label Policy

Overall, firms appear to respond to differences in the regulatory costs of entry by forum shopping and relying on off-label drug use. Many off-label uses are not supported by high-quality scientific evidence, yet at the same time, such use represents an important source of medical innovation, offering earlier access to potential treatments for patients who may not respond to approved drugs (Radley et al., 2006; Stafford, 2008). Although FDA guidance on off-label promotion has gradually loosened over time (FDA, 2014), policymakers and regulators have recently proposed legal provisions that would ban certain off-label uses (Zinberg, 2023).

While we leave to future research a full welfare analysis of the costs and benefits of banning off-label use, we consider its implications from a conceptual standpoint.[26] Our calculations above on the impact of regulatory slack on firms' market entry investments suggest a sizeable portion of investment goes toward indications used off-label. Hence, regulators considering whether to limit such use must weigh any gains in information quality for once off-label indications that must now go through regulatory approval processes against the loss in potential therapeutic options due to missing or delayed indications.

6 Conclusion

The high costs and risks of regulation necessitate an understanding of how firms navigate regulatory processes when introducing new products. Combining data on cancer drug approvals with information from large-scale genomic sequencing efforts, we measure the potential shoppability of new technologies and document how their manufacturers selectively engage with more favorable regulatory environments. Our results suggest that pharmaceutical firms forum shop by seeking regulatory approval in smaller indications and exploiting slack in regulatory requirements (i.e., the possibility of off-label drug use) to increase demand for their drugs outside of formal approval

[26] For recent work on this, see Tunçel (forthcoming).

processes. Manufacturers of more shoppable drugs, with higher expected off-label use, are more likely to forum shop. Additional analyses, leveraging variation in the likelihood of off-label use across therapeutic markets and differences in drug and firm types, further support this view.

These findings have important implications. For managers, our results demonstrate the importance of forum shopping: Firms can lower the costs of market entry by initially targeting smaller indications, which are associated with shorter development times, rather than pursuing longer and costlier approvals in larger markets.[27] We calculate that by seeking initial approval in a smaller market, firms can reduce their drug development timelines by an average of 7.9 months. This reduction in development time corresponds to a value of $5.3 million, plus an additional $94.3 million in revenues from earlier market entry. Forum shopping for initial entry markets may present a viable strategy in other regulated industries as well, as firms introducing new products in these industries often face high costs associated with meeting regulatory requirements and benefit from first-mover advantages.[28]

Our analysis contains a few limitations and suggests additional opportunities for future research. First, our study primarily focuses on how differences in regulatory costs across markets may influence firms to forum shop. However, other factors, such as prices, may also play an important role in market entry decisions. For example, firms might seek initial approval for rare diseases, which may allow them to set higher prices due to relatively inelastic demand among the initial patient population. These high prices can then be used for any future uses of the product, both approved and off-label (Chandra and Garthwaite, 2017). We hope that our findings serve as a foundation for further investigations into the role of entry regulation on firms' market entry decisions and forum shopping practices that incorporate data on key factors, such as prices. Second, because we cannot directly observe regulatory investments that do not lead to success (i.e., NDAs and sNDAs that are not approved), our estimates may not fully capture firms' market entry decisions. Nonetheless, this concern is mitigated by the fact that firms generally do not submit applications to the FDA that they anticipate will be rejected. Finally, by accounting for variation in the shoppability of

[27] In these situations, firms may benefit by getting a drug into the hands of physicians and patients as quickly as possible via one indication approval and then relying on off-label use or later regulatory approvals for broader indications.

[28] For example, an automobile manufacturer may develop various prototypes aimed at different market segments but initially launch the model most likely to meet vehicle emissions standards (Pinkse et al., 2014).

drugs, we extend existing studies which typically consider the decision to forum shop as fixed across products.[29] However, the degree to which technologies are shoppable may vary by industry, highlighting the need for future work examining other contexts.

A full welfare analysis of the role of entry regulation on firms' market entry decisions is beyond the scope of this paper. However, our findings point to the need to think deeply about the costs of regulatory approval and corresponding policies. Many drugs are recommended and used off-label for important health conditions, but off-label use without sufficient evidence is also associated with higher rates of adverse events (Eguale et al., 2016; Richardson, 2016; Wittich et al., 2012). Our analysis thus highlights the critical trade-off that regulators must navigate between encouraging expedient access to drugs and the need for sufficient quality information on new therapies. At one extreme, stringent regulatory processes may slow the entry of valuable products. At the other, firms may choose to avoid regulatory approval entirely, leading to a dearth of valuable new products and insufficient information regarding the quality of those that bypass formal regulatory approval.

[29] Sytch and Kim (2021) also explore heterogeneity in firms' ability to forum shop, noting the importance of shared social connections between lawyers representing firms and judges.

References

Acemoglu, D. and J. Linn (2004). Market size in innovation: Theory and evidence from the pharmaceutical industry. *Quarterly Journal of Economics* 119(3): 1049–1090.

Adams, C. P. and V. V. Brantner (2006). Estimating the cost of new drug development: Is it really $802 million? *Health Affairs* 25(2): 420–428.

Avorn, J., A. Sarpatwari, and A. S. Kesselheim (2015). Forbidden and permitted statements about medications—Loosening the rules. *New England Journal of Medicine* 373(10): 967–973.

Bagley, N., A. Chandra, C. Garthwaite, and A. D. Stern (2018). It's time to reform the Orphan Drug Act. *NEJM Catalyst* 4(6).

Bebchuk, L. A. and A. Cohen (2003). Firms' decisions where to incorporate. *Journal of Law and Economics* 46(2): 383–425.

Blume-Kohout, M. E. and N. Sood (2013). Market size and innovation: Effects of Medicare Part D on pharmaceutical research and development. *Journal of Public Economics* 97: 327–336.

Bothwell, L. E., J. A. Greene, S. H. Podolsky, and D. S. Jones (2016). Assessing the gold standard—Lessons from the history of RCTs.

Bradford, W. D., J. L. Turner, and J. W. Williams (2018). Off-label use of pharmaceuticals: A detection controlled estimation approach. *Journal of Industrial Economics* 66(4): 866–903.

Buchak, G., G. Matvos, T. Piskorski, and A. Seru (2018). Fintech, regulatory arbitrage, and the rise of shadow banks. *Journal of Financial Economics* 130(3): 453–483.

Budish, E., B. N. Roin, and H. Williams (2015). Do firms underinvest in long-term research? Evidence from cancer clinical trials. *American Economic Review* 105(7): 2044–2085.

Cain, M. D. and S. M. Davidoff (2012). Delaware's competitive reach. *Journal of Empirical Legal Studies* 9(1): 92–128.

Chandra, A. and C. Garthwaite (2017). The economics of indication-based drug pricing. *New England Journal of Medicine* 377(2): 103–106.

Chandra, A., C. Garthwaite, and A. D. Stern (2019). Characterizing the drug development pipeline for precision medicines. In E. R. Berndt, D. P. Goldman, and J. W. Rowe (Eds.), *Economic Dimensions of Personalized and Precision Medicine*, pp. 115–157. Chicago: University of Chicago Press.

Cheng, F., W. Lu, C. Liu, J. Fang, Y. Hou, D. E. Handy, R. Wang, Y. Zhao, Y. Yang, J. Huang, D. E. Hill, M. Vidal, C. Eng, and J. Loscalzo (2019). A genome-wide positioning systems network algorithm for in silico drug repurposing. *Nature Communications* 10(1): 3476.

Cohen, L., U. G. Gurun, and S. D. Kominers (2019). Patent trolls: Evidence from targeted firms. *Management Science* 65(12): 5461–5486.

Cook, D., D. Brown, R. Alexander, R. March, P. Morgan, G. Satterthwaite, and M. N. Pangalos (2014). Lessons learned from the fate of AstraZeneca's drug pipeline: a five-dimensional framework. *Nature Reviews Drug Discovery* 13(6): 419–431.

CSDD (2014). Costs to develop and win marketing approval for a new drug is $2.6 billion. `http://csdd.tufts.edu/news/complete_story/pr_tufts_csdd_2014_cost_study`, Tufts Center for the Study of Drug Development.

Danzon, P. M. and E. L. Keuffel (2014). Regulation of the pharmaceutical-biotechnology industry. In N. L. Rose (Ed.), *Economic Regulation and Its Reform: What Have We Learned?*, pp. 407–484. Chicago: University of Chicago Press.

DiMasi, J. A. (2001). New drug development in the United States from 1963 to 1999. *Clinical Pharmacology & Therapeutics* 69(5): 286–296.

DiMasi, J. A., R. W. Hansen, and H. G. Grabowski (2003). The price of innovation: New estimates of drug development costs. *Journal of Health Economics* 22(2): 151–185.

Douglas, J. L. (2016). New wine into old bottles: Fintech meets the bank regulatory world. *North Carolina Banking Institute* 20**:** 17.

Dubois, P., E. Gentry, and T. Tunçel (2023). Strategic submissions: A cross-country analysis of supplemental drug approvals. TSE Working Paper, n. 23-1459.

Eguale, T., D. L. Buckeridge, A. Verma, N. E. Winslade, A. Benedetti, J. A. Hanley, and R. Tamblyn (2016). Association of off-label drug use and adverse drug events in an adult population. *JAMA Internal Medicine* 176(1)**:** 55–63.

Ellias, J. A. (2018). What drives bankruptcy forum shopping? Evidence from market data. *Journal of Legal Studies* 47(1)**:** 119–149.

FDA (1998). Guidance for industry: FDA approval of new cancer treatment uses for marketed and biological products.

FDA (2014). Guidance for industry: Distributing scientific and medical publications on unapproved new uses—recommended practices.

FDA (2024a). The drug development process—Step 3: Clinical research. `https://www.fda.gov/patients/drug-development-process/step-3-clinical-research`, accessed on 2024-05-04.

FDA (2024b). Office of Orphan Products Development (OOPD) database. `https://www.accessdata.fda.gov/scripts/opdlisting/oopd/`, accessed on 2024-05-04.

Gelijns, A. C., N. Rosenberg, and A. J. Moskowitz (1998). Capturing the unexpected benefits of medical research. *New England Journal of Medicine* 339(10)**:** 693–698.

Hee, S. W., A. Willis, C. Tudur Smith, S. Day, F. Miller, J. Madan, M. Posch, S. Zohar, and N. Stallard (2017). Does the low prevalence affect the sample size of interventional clinical trials of rare diseases? An analysis of data from the aggregate analysis of clinicaltrials.gov. *Orphanet Journal of Rare Diseases* 12**:** 1–20.

Holmes, T. J. (1998). The effect of state policies on the location of manufacturing: Evidence from state borders. *Journal of Political Economy* 106(4)**:** 667–705.

Houston, J. F., C. Lin, and Y. Ma (2012). Regulatory arbitrage and international bank flows. *Journal of Finance* 67(5)**:** 1845–1895.

IQVIA (2023). Success multiplied: Launch excellence for multi-indication assets. `https://www.iqvia.com/library/white-papers/success-multiplied-launch-excellence-for-multi-indication-assets`, accessed on 2024-05-07.

IQVIA Institute (2018). Global oncology trends 2018. `https://www.iqvia.com/insights/the-iqvia-institute/reports/global-oncology-trends-2018`, accessed on 2021-01-13.

Joskow, P. L. (2005). Regulation and deregulation after 25 years: Lessons learned for research in industrial organization. *Review of Industrial Organization* 26**:** 169–193.

Kakani, P., M. Chernew, and A. Chandra (2022). The contribution of price growth to pharmaceutical revenue growth in the United States: Evidence from medicines sold in retail pharmacies. *Journal of Health Politics, Policy and Law* 47(6)**:** 629–648.

Karolyi, G. A. and A. G. Taboada (2015). Regulatory arbitrage and cross-border bank acquisitions. *Journal of Finance* 70(6)**:** 2395–2450.

Krieger, J., D. Li, and D. Papanikolaou (2022). Missing novelty in drug development. *Review of Financial Studies* 35(2)**:** 636–679.

Krieger, J. L. (2021). Trials and terminations: Learning from competitors' R&D failures. *Management Science* 67(9)**:** 5525–5548.

Larkin, I., D. Ang, J. Avorn, and A. S. Kesselheim (2014). Restrictions on pharmaceutical detailing reduced off-label prescribing of antidepressants and antipsychotics in children. *Health Affairs* 33(6)**:** 1014–1023.

Lee, E., A. R. Teschemaker, R. Johann-Liang, G. Bazemore, M. Yoon, K.-S. Shim, M. Daniel, J. Pittman, and A. K. Wutoh (2012). Off-label prescribing patterns of antidepressants in children and adolescents. *Pharmacoepidemiology and Drug Safety* 21(2)**:** 137–144.

Lerner, J. and J. Tirole (2006). A model of forum shopping. *American Economic Review* 96(4): 1091–1113.

Maini, L. and F. Pammolli (2023). Reference pricing as a deterrent to entry: Evidence from the European pharmaceutical market. *American Economic Journal: Microeconomics* 15(2): 345–383.

Martin, L., M. Hutchens, C. Hawkins, and A. Radnov (2017). How much do clinical trials cost? *Nature Reviews Drug Discovery* 16(6): 381–382.

McKibbin, R. (2023). The effect of RCTs on drug demand: Evidence from off-label cancer drugs. *Journal of Health Economics* 90: 102779.

Melsheimer, R., A. Geldhof, I. Apaolaza, and T. Schaible (2019). Remicade® (infliximab): 20 years of contributions to science and medicine. *Biologics: Targets and Therapy*: 139–178.

Michaeli, T., H. Jürges, and D. T. Michaeli (2023). FDA approval, clinical trial evidence, efficacy, epidemiology, and price for non-orphan and ultra-rare, rare, and common orphan cancer drug indications: cross sectional analysis. *BMJ* 381: e073242.

Miller, K. L. and M. Lanthier (2018). Investigating the landscape of US orphan product approvals. *Orphanet Journal of Rare Diseases* 13: 1–8.

Mullard, A. (2016). Parsing clinical success rates. *Nature Reviews Drug Discovery* 15(7): 447–448.

NIH (2024). Autoimmune diseases of rheumatoid arthritis and lupus. National Institutes of Health Accelerating Medicines Partnership (AMP). `https://www.nih.gov/research-training/accelerating-medicines-partnership-amp/autoimmune-diseases-rheumatoid-arthritis-lupus`, accessed on 2024-05-04.

Pfister, D. G. (2012). Off-label use of oncology drugs: The need for more data and then some. *Journal of Clinical Oncology* 30(6): 584–586.

Pinkse, J., R. Bohnsack, and A. Kolk (2014). The role of public and private protection in disruptive innovation: The automotive industry and the emergence of low-emission vehicles. *Journal of Product Innovation Management* 31(1): 43–60.

Pleasance, E., A. Bohm, L. Williamson, J. Nelson, Y. Shen, M. Bonakdar, E. Titmuss, V. Csizmok, K. Wee, S. Hosseinzadeh, and Others (2022). Whole-genome and transcriptome analysis enhances precision cancer treatment options. *Annals of Oncology* 33(9): 939–949.

Posen, H. A. (2015). Ridesharing in the sharing economy: Should regulators impose Über regulations on Uber. *Iowa Law Review* 101: 405.

Radley, D. C., S. N. Finkelstein, and R. S. Stafford (2006). Off-label prescribing among office-based physicians. *Archives of Internal Medicine* 166(9): 1021–1026.

Richardson, E. (2016). Off-label drug promotion. *Health Affairs* Health Policy Brief, DOI: `10.1377/hpb20160630.920075`.

Schuhmacher, A., M. Hinder, N. Boger, D. Hartl, and O. Gassmann (2022). The significance of blockbusters in the pharmaceutical industry. *Nature Reviews Drug Discovery*.

Seoane-Vazquez, E., R. Rodriguez-Monguio, S. L. Szeinbach, and J. Visaria (2008). Incentives for orphan drug research and development in the United States. *Orphanet Journal of Rare Diseases* 3(1): 1–7.

Sertkaya, A., H.-H. Wong, A. Jessup, and T. Beleche (2016). Key cost drivers of pharmaceutical clinical trials in the United States. *Clinical Trials* 13(2): 117–126.

Shapiro, B. T. (2018). Informational shocks, off-label prescribing, and the effects of physician detailing. *Management Science* 64(12): 5925–5945.

Sondka, Z., S. Bamford, C. G. Cole, S. A. Ward, I. Dunham, and S. A. Forbes (2018). The COSMIC Cancer Gene Census: Describing genetic dysfunction across all human cancers. *Nature Reviews Cancer* 18(11): 696–705.

Stafford, R. S. (2008). Regulating off-label drug use—Rethinking the role of the FDA. *New England Journal of Medicine* 358(14): 1427–1429.

Sytch, M. and Y. H. Kim (2021). Quo vadis? From the schoolyard to the courtroom. *Administrative Science Quarterly* 66(1): 177–219.

Tanoli, Z., M. Vähä-Koskela, and T. Aittokallio (2021). Artificial intelligence, machine learning, and drug repurposing in cancer. *Expert Opinion on Drug Discovery* 16(9)**:** 977–989.

Tate, J. G., S. Bamford, H. C. Jubb, Z. Sandra, D. M. Beard, N. Bindal, H. Boutselakis, C. G. Cole, C. Creatore, E. Dawson, P. Fish, B. Harsha, C. Hathaway, S. C. Jute, C. Y. Kok, K. Noble, L. Ponting, C. C. Ramshaw, C. E. Rye, H. E. Speedy, R. Stefancsik, S. L. Thompson, S. Wang, S. Ward, P. J. Campbell, and S. A. Forbes (2018). COSMIC: The catalogue of somatic mutations. *Nucleic Acids Research* 47(D1)**:** D941–47.

TCGA Research Network (2011). Integrated genomic analyses of ovarian carcinoma. *Nature* 474(7353)**:** 609–615.

Tunçel, T. (forthcoming). Should we prevent off-label drug prescriptions? Empirical evidence from France. *Review of Economic Studies*.

Uffelmann, E., Q. Q. Huang, N. S. Munung, J. De Vries, Y. Okada, A. R. Martin, H. C. Martin, T. Lappalainen, and D. Posthuma (2021). Genome-wide association studies. *Nature Reviews Methods Primers* 1(1)**:** 59.

Vives, X. (2019). Digital disruption in banking. *Annual Review of Financial Economics* 11**:** 243–272.

Weinstein, J. N., E. A. Collisson, G. B. Mills, K. R. Shaw, B. A. Ozenberger, K. Ellrott, I. Shmulevich, C. Sander, and J. M. Stuart (2013). The Cancer Genome Atlas Pan-Cancer analysis project. *Nature Genetics* 45(10)**:** 1113–1120.

Wittich, C. M., C. M. Burkle, and W. L. Lanier (2012). Ten common questions (and their answers) about off-label drug use. *Mayo Clinic Proceedings* 87(10)**:** 982–990.

Wong, H.-H., A. Jessup, A. Sertkaya, A. Birkenbach, A. Berlind, and J. Eyraud (2014). Examination of clinical trial costs and barriers for drug development final. Office of the Assistant Secretary for Planning and Evaluation, US Department of Health & Human Services.

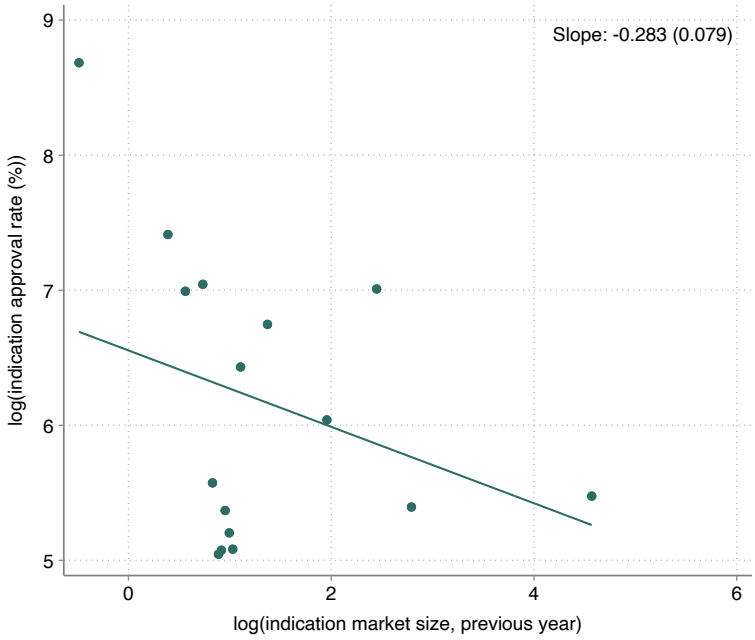
Wouters, O. J., M. McKee, and J. Luyten (2020). Estimated research and development investment needed to bring a new medicine to market, 2009-2018. *JAMA* 323(9)**:** 844–853.

Zinberg, J. (2023). The FDA wants to interfere in the practice of medicine. `https://www.wsj.com/articles/the-fda-wants-to-interfere-in-the-practice-of-medicine-physicians-patients-medical-devices-treatment-11673562165`, *Wall Street Journal*, January 12, 2023.

Online Appendices

A Approval Rates Across Indications

FIGURE A1: APPROVAL RATES ACROSS INDICATIONS



NOTES: This figure shows a binned scatter plot comparing indication approval rates and market size for cancer sites, for cancer drugs approved from 1978 to 2016. The level of observation is the cancer site-year. Indication approval rates are defined as the probability that a drug tested in pre-clinical research in a given year advances to approval. The figure compares the log of the indication approval rate to the log of the market size in the year prior to pre-clinical testing.

B Similarity Index Example: Breast and Ovarian Cancers

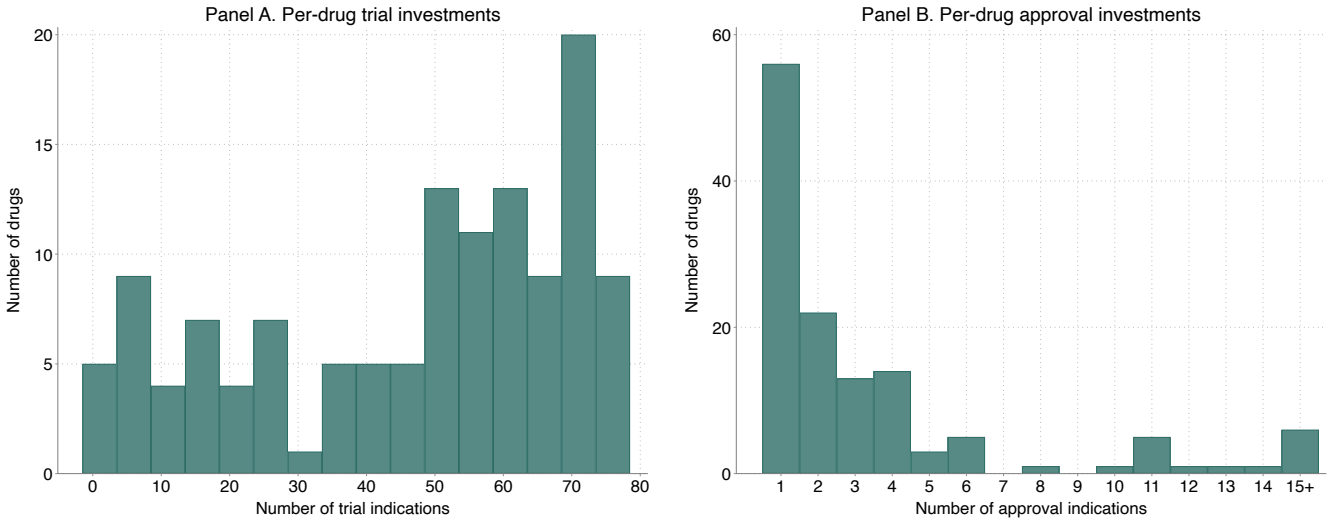
Table B1 provides the genetic mutations associated with breast and ovarian cancers from the CGC database. We calculate their similarity index as follows: $s_{Breast,Ovarian} = \frac{|B \cap O|}{|B \cup O|} = 10/85 = 0.118$.

Table B1: Genetic Mutations for Breast and Ovarian Cancers

Breast gene mutations	**Ovarian gene mutations**	**Breast & ovarian gene mutations**
ALK	AKT2	AKT1
APOBEC3B	ATR	ARID1A
ASPM	BRAF	ARID1B
BAP1	CASP3	BARD1
BRIP1	CCNE1	BRCA1
CASP8	CDK12	BRCA2
CCND1	COL3A1	ERBB2
CDH1	CREB1	GOLPH3
CDKN1B	CSMD3	PPM1D
CHEK2	CTNNB1	RAD50
CTCF	EIF1AX	
DCTN1	EWSR1	
EP300	FES	
ESR1	FOXL2	
ETV6	GOPC	
FADD	LRP1B	
FBLN2	MAPK1	
FEN1	MLH1	
FLCN	MSH2	
FLNA	MSH6	
FOXA1	PIK3R1	
GATA3	PLAG1	
HGF	PMS1	
IKZF3	PMS2	
IRS4	PPP2R1A	
MAP2K4	PRDM2	
MAP3K1	PTK6	
MAP3K13	RNF43	
MED12	ROS1	
NCOR1	STK11	
NOTCH1		
NTRK3		
PALB2		
PBRM1		
PIK3CA		
POLQ		
PPFIBP1		
RANBP2		
RB1		
SALL4		
SMARCD1		
TBX3		
TP53		
VHL		
ZMYM3		

C Variation in Approval and Trial Indications

FIGURE C1: PER-DRUG DISTRIBUTION OF MARKET ENTRY INVESTMENTS



NOTES: This figure shows the distribution of market entry investments for cancer drugs approved from 1978 to 2016. Panel A gives the per-drug distribution of trial indications; number of drugs is 127. Panel B gives the per-drug distribution of approval indications; number of drugs is 129.

D Robustness Checks

D.1 Difference Between First and Subsequent Indications

TABLE D1: ORDERING OF MARKET ENTRY INVESTMENTS

	Clinical trial investments		Regulatory approval investments	
	(1)	(2)	(3)	(4)
$\mathbb{1}_{First\ indication}$	1.073***	0.219	-0.386*	-0.269
	(0.122)	(0.151)	(0.220)	(0.208)
Mean of dep. var	6.825	6.794	6.513	6.522
Observations	2,007	1,783	187	182
Controls:				
Initial approval year	no	yes	no	yes
Indication group	no	yes	no	yes
Competition	no	yes	no	yes
Regulatory incentives	no	yes	no	yes
Intellectual property	no	yes	no	yes

NOTES: This table shows the difference in market size between the first and subsequent indications for cancer drugs approved from 1978 to 2016. The level of observation is the drug-indication order. Columns 1 and 2 examine clinical trial investments, and Columns 3 and 4 examine regulatory approval investments. The outcome variable is market size, measured by new diagnoses for an indication in the SEER data. We use the log of the five-year average market size associated with indication order. The number of observations differs between Columns 1 and 2 (and between Columns 3 and 4) due to missing data on control variables and dropped singletons. Robust standard errors are in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

TABLE D2: ORDERING OF MARKET ENTRY INVESTMENTS, INCORPORATING OFF-LABEL OPPORTUNITIES

	Clinical trial investments		Regulatory approval investments	
	Focal market size (1)	Total market size (2)	Focal market size (3)	Total market size (4)
$\mathbb{1}_{First\ indication}$	0.219	0.180	-0.269	0.315*
	(0.151)	(0.149)	(0.208)	(0.174)
Mean of dep. var	6.794	7.966	6.522	8.981
Observations	1,783	1,783	182	182
Controls:				
Initial approval year	yes	yes	yes	yes
Indication group	yes	yes	yes	yes
Competition	yes	yes	yes	yes
Regulatory incentives	yes	yes	yes	yes
Intellectual property	yes	yes	yes	yes

NOTES: This table shows the difference in market size between the first and subsequent indications for cancer drugs approved from 1978 to 2016. The level of observation is the drug-indication order. Columns 1 and 2 examine clinical trial investments, and Columns 3 and 4 examine regulatory approval investments. The outcome variable in Columns 1 and 3 is focal market size, while that in Columns 2 and 4 is total market size, including potential off-label markets. Focal market size is measured by new diagnoses for an indication in the SEER data. Total market size is measured by new diagnoses for the focal indication plus a proportion of new diagnoses for any potential off-label indications, with the proportions given by our disease similarity index. For both variables, we use the log of the five-year average market size associated with indication order. Robust standard errors are in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

D.2 Cross-Disease Analysis

D.2.1 Controlling for Drug and Market Characteristics

TABLE D3: ORDERING OF REGULATORY APPROVAL INVESTMENTS ACROSS DISEASES, BY PROPENSITY FOR OFF-LABEL DRUG USE

	Low off-label diseases		High off-label diseases	
	(1)	(2)	(3)	(4)
Indication order	-2.202***	-0.840***	-0.265***	0.0362
	(0.277)	(0.245)	(0.0617)	(0.0684)
Mean of Dep. Var.	16.044	16.160	13.524	13.540
Observations	152	123	650	571
Controls:				
Initial approval year	no	yes	no	yes
Indication group	no	yes	no	yes
Competition	no	yes	no	yes
Regulatory incentives	no	yes	no	yes

NOTES: This table shows the relationship between indication order for regulatory approval investments and market size for approved drugs across several disease categories from 1998 to 2021, by propensity for off-label drug use. The level of observation is the drug-indication order. The outcome variable is market size, measured by new diagnoses for an indication (ICD-9) in the MEPS data. We use the log of the market size associated with focal indication order averaged over 1996 and 1997. The number of observations differs between Columns 1 and 2 (and between Columns 3 and 4) due to missing data on control variables and dropped singletons. Robust standard errors are in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

D.2.2 Controlling for ATC Classification

TABLE D4: ORDERING OF REGULATORY APPROVAL INVESTMENTS ACROSS DISEASES, BY PROPENSITY FOR OFF-LABEL DRUG USE, WITH ALTERNATIVE DISEASE CONTROLS

	Low off-label diseases (1)	High off-label diseases (2)
Indication order	-0.978***	0.0365
	(0.285)	(0.0688)
Mean of Dep. Var.	16.164	13.529
Observations	119	566
Controls:		
Initial approval year	yes	yes
Indication group	yes	yes
Competition	yes	yes
Regulatory incentives	yes	yes

NOTES: This table shows the relationship between indication order for regulatory approval investments and market size for approved drugs across several disease categories from 1998 to 2021, controlling for each drug's ATC classification. There are 14 mutually exclusive ATC categories: alimentary tract and metabolism, anti-infectives for systemic use, antineoplastic and immunomodulating agents, antiparasitic products, insecticides and repellents, blood and blood forming clots, cardiovascular system, dermatologicals, genitourinary system and sex hormones, musculoskeletal system, nervous system, respiratory system, sensory organs, systemic hormonal preparations, and various. The level of observation is the drug-indication order. The outcome variable is market size, measured by new diagnoses for an indication (ICD-9) in the MEPS data. We use the log of the market size associated with focal indication order averaged over 1996 and 1997. The number of observations differs between this table and Table D3 due to dropped singletons. Robust standard errors are in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

D.3 Incorporating an Alternative Measure of Actual Off-Label Drug Use

We incorporate into the cross-disease analysis an alternative measure of actual off-label drug use using physician prescribing data from the IMS Health National Disease and Therapeutic Index^{TM} (NDTI^{TM}) from 2004 to 2009. The NDTI^{TM} provides data based on surveys of practicing physicians, allowing us to measure both on-label and off-label drug use. For each drug, we calculate the share of prescriptions that are off-label in a given year. To examine how off-label use shapes the indication order-market size correlation, we restrict our analysis to drug approvals for a given period and control for the share of off-label prescriptions in the year prior to the start of the analysis period. For example, Panel A in Table D5 focuses on the set of all drugs approvals starting in 2005. Column 1 shows the correlation between indication order and market size. Column 2 controls for the share of off-label prescriptions in 2004 and shows that the indication order-market size correlation is more negative after accounting for off-label drug use. We observe a similar pattern in subsamples beginning in subsequent years (2006, 2007, 2008, 2009, and 2010).

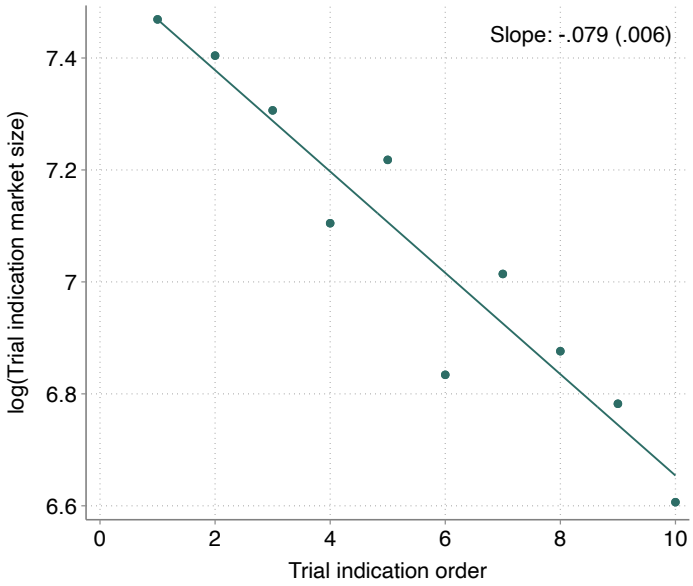
TABLE D5: ORDERING OF REGULATORY APPROVAL INVESTMENTS ACROSS DISEASES, ACCOUNTING FOR ACTUAL OFF-LABEL DRUG USE

	(1)	(2)
Panel A. Drugs approved 2005+		
Indication order	-0.312***	-0.325***
	(0.0583)	(0.0594)
Mean of Dep. Var.	13.838	13.838
Observations	716	716
Panel B. Drugs approved 2006+		
Indication order	-0.308***	-0.332***
	(0.0591)	(0.0598)
Mean of Dep. Var.	13.796	13.796
Observations	680	680
Panel C. Drugs approved 2007+		
Indication order	-0.296***	-0.323***
	13.766	13.766
Mean of Dep. Var.	(0.0589)	(0.0595)
Observations	650	650
Panel D. Drugs approved 2008+		
Indication order	-0.317***	-0.337***
	(0.0610)	(0.0615)
Mean of Dep. Var.	13.724	13.724
Observations	618	618
Panel E. Drugs approved 2009+		
Indication order	-0.322***	-0.337***
	(0.0620)	(0.0623)
Mean of Dep. Var.	13.694	13.694
Observations	584	584
Panel F. Drugs approved 2010+		
Indication order	-0.297***	-0.318***
	(0.0614)	(0.0622)
Mean of Dep. Var.	13.619	13.619
Observations	547	547

NOTES: This table shows the relationship between indication order for regulatory approval investments and market size for approved drugs across several disease categories. Each coefficient is from a separate regression. Regressions in Column 2 control for the share of off-label prescriptions in the year prior to the start of the analysis period, while those in Column 1 do not. The outcome variable is market size, measured by new diagnoses for an indication (ICD-9) in the MEPS data. We use the log of the market size associated with focal indication order averaged over 1996 and 1997. Robust standard errors are in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

D.4 Trial End Dates

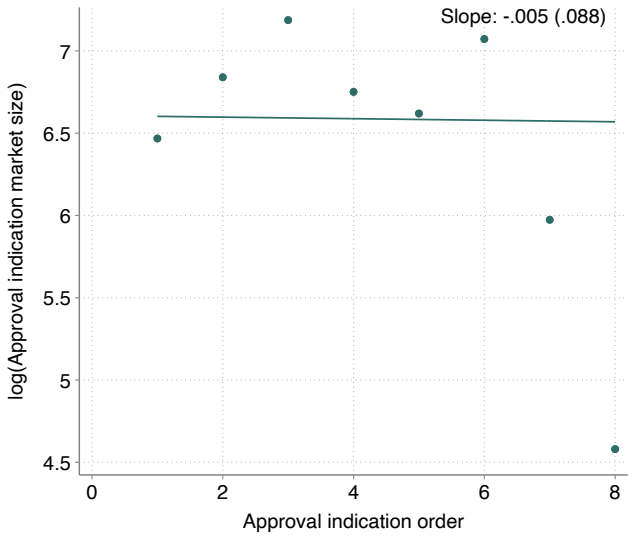
FIGURE D1: TRENDS IN CLINICAL TRIAL INVESTMENTS, WITH INDICATION ORDER DETERMINED BY TRIAL END DATES



NOTES: This figure shows trends in clinical trial investments for cancer drugs approved from 1978 to 2016, with indication order determined by trial end dates. The level of observation is the drug-indication order. Market size is measured by new diagnoses for an indication in the SEER data. We use the log of the five-year average market size. Each marker represents binned averages for a given indication order. For ease of intrepretation, we display up to the 10th indication.

D.5 Submission Dates

FIGURE D2: TRENDS IN REGULATORY INVESTMENTS, WITH INDICATION ORDER DETERMINED BY SUBMISSION DATES



NOTES: This figure shows trends in regulatory approval investments for cancer drugs approved from 1978 to 2016, with indication order determined by FDA submission dates. The level of observation is the drug-indication order. Market size is measured by new diagnoses for an indication in the SEER data. We use the log of the five-year average market size. Each marker represents binned averages for a given indication order.